Some problems of statistics and optimal control for stochastic processes in the field of electricity markets prices modeling

Pierre Gruet

Université Paris Diderot, Sorbonne Paris Cité

École doctorale ED386 Sciences Mathématiques de Paris Centre

Laboratoire de Probabilités et Modèles Aléatoires

THÈSE DE DOCTORAT

Discipline : Mathématiques appliquées

Présentée par

Pierre Gruet

Quelques problèmes d'estimation et de contrôle optimal pour les processus stochastiques dans un cadre de modélisation des prix des marchés de l'électricité

Some problems of statistics and optimal control for stochastic processes in the field of electricity markets prices modeling

---

Sous la direction de Huyên Pham
et la codirection de Marc Hoffmann

Soutenue publiquement le 2 décembre 2015 devant le jury composé de

René Aïd	*EDF R&D*	Examinateur
Luciano Campi	*London School of Economics*	Rapporteur
Olivier Féron	*EDF R&D*	Examinateur
Arnaud Gloter	*Université d'Évry-Val-d'Essonne*	Rapporteur
Marc Hoffmann	*Université Paris Dauphine*	Codirecteur de thèse
Huyên Pham	*Université Paris Diderot*	Directeur de thèse
Mathieu Rosenbaum	*Université Pierre et Marie Curie*	Examinateur

Remerciements

Cette thèse de doctorat s'ouvre par la dernière contribution que j'y apporte : voici le moment de remercier tous ceux qui ont accompagné l'épreuve d'endurance que celle-ci constitue.

Mes premières pensées vont à Huyên Pham et à Marc Hoffmann, qui ont encadré mes travaux et m'ont fait découvrir la recherche académique. Marc m'a montré la richesse de la statistique des processus et me parlait de la recherche avec des métaphores montagnardes et guerrières. Huyên m'a initié au contrôle stochastique et m'a permis de suivre plusieurs conférences parfois lointaines. Tous deux se sont toujours montrés disponibles tout autant qu'exigeants. J'ai une grande gratitude envers eux.

René Aïd et Olivier Féron ont été des collaborateurs très présents pendant ces trois années. Je les remercie pour leurs explications quant au contexte d'application de mes travaux et pour les orientations qu'ils ont contribué à leur donner, mais aussi pour leur sympathie et leur humour qui m'ont fait passer de très bons moments que je garderai en mémoire.

J'adresse ensuite de sincères remerciements à Arnaud Gloter et Luciano Campi ; je suis très honoré qu'ils aient accepté de rapporter mes travaux de thèse et qu'ils les aient lus avec tant d'attention.

Merci aussi à Mathieu Rosenbaum, qui complète le jury après avoir suivi l'évolution de ma recherche d'assez près.

Tout au long de ces trois années, j'ai bénéficié d'un très bon environnement scientifique au sein du LPMA et du laboratoire de Finance des Marchés de l'Énergie (FiME). Que leurs membres soient ici remerciés pour leurs conseils et leurs remarques au cours de nos discussions et de divers exposés. J'ai également une pensée pour Frédéric Hélein et Olivier Bokanowski, avec qui j'ai effectué des enseignements. Côté administratif, l'efficacité de Nathalie Bergame, Pascal Chiettini et Valérie Juvé a été très appréciable.

Je suis reconnaissant à la très compétente équipe d'informaticiens de l'UFR de mathématiques, ainsi qu'aux membres du projet Debian, qui m'ont permis de disposer d'un excellent environnement informatique pendant la préparation de ma thèse.

Je remercie aussi les membres du groupe R32 du département OSIRIS, chez EDF R&D, pour leur sympathie et leurs encouragements.

Enfin, une partie de cette thèse a été rédigée à l'*Institute for Pure and Applied Mathematics*, à Los Angeles. J'y ai trouvé un excellent cadre de travail ainsi qu'un personnel très aimable, attentif et efficace.

Au quotidien, la présence de collègues doctorants offre un réconfort certain ; au LPMA, j'ai apprécié de côtoyer Adrien, Anna, Arturo, Aser, Christophe, Clément, David, Guillaume, Huy, Jiatu, Lorick, Marc-Antoine, Maud, Nicolas, Oriane, Pietro, Shanqiu, Sophie, Thomas G., Thomas V. et Vu-Lan. Chez EDF R&D, j'ai connu plus spécifiquement Erwan, Thomas D., Vincent et Youcef.

Quelques personnes ont occupé une place spéciale durant la préparation de ma thèse ; chez EDF, j'ai partagé mon bureau, mes pauses et de grandes discussions musicales ou culinaires avec Nedjmeddine Allab. Nous avons passé beaucoup de temps ensemble et avons préparé nos thèses respectives sans nous départir d'un humour toujours vif.

Lors de mes passages à l'université Paris Dauphine, j'ai passé plusieurs journées avec Damien Fessler, qui m'a prodigué de bons conseils, tant en matière de recherche académique que de musique classique.

Dès ma scolarité en Master 2 et à l'ENSAE, j'ai connu Adélaïde Olivier, qui a commencé sa thèse en même temps que moi. Nous avons partagé un grand nombre de déjeuners à Jussieu et avons maintes fois refait le monde. Je suis heureux de me prévaloir de son amitié.

À côté de cette thèse, j'ai pris un grand plaisir à pratiquer le tir à l'arc, ainsi qu'à entraîner des archers et à administrer l'activité. J'ai notamment beaucoup apprécié de travailler avec mes collègues des conseils d'administration de la Compagnie de Châtillon et du Comité départemental des Hauts-de-Seine, ainsi qu'avec nos formateurs.

Par ailleurs, j'ai beaucoup apprécié la présence de Didier et Antoine, avec qui nous avons mis en place d'héroïques journées de travail commençant bien avant le lever du soleil. J'ai passé de bons moments avec mes amis et voisins Fanny et Arnaud. Les derniers jours de rédaction de cette thèse ont été passés chez Caroline, Yannik et leur fille Lara, qui m'a emmené faire du toboggan pour prendre l'air. Je remercie chaleureusement mon amie Florence pour son soutien, ses cartes postales et les restaurants qu'elle m'a fait découvrir.

Pour finir, merci à ma famille, qui m'a encouragé et soutenu tout en apportant son aide logistique *via* de solides repas et en m'emmenant faire de belles promenades, au cours desquelles nous avons pratiqué le *geocaching*. Merci à Miranda pour son attention de tous les instants, sa bonne humeur permanente, sa relecture attentive et nos discussions de statisticiens. Elle a largement participé à l'épreuve d'endurance...

Résumé

Cette thèse porte sur l'étude de modèles mathématiques de l'évolution des prix sur les marchés de l'électricité, du point de vue de la statistique des processus et de celui du contrôle optimal stochastique.

Dans une première partie, nous estimons les composantes de la volatilité d'un processus de diffusion multidimensionnel représentant l'évolution des prix sur le marché à terme de l'électricité. Sa dynamique est conduite par deux mouvements browniens. Nous cherchons à réaliser l'estimation efficacement en termes de vitesse de convergence, et de variance limite en ce qui concerne la partie paramétrique de ces composantes. Cela nécessite une extension de la définition usuelle de l'efficacité au sens de Cramér-Rao. Nos méthodes d'estimation sont fondées sur la variation quadratique réalisée du processus observé.

Dans la deuxième partie, nous ajoutons des termes d'erreur de modèle aux observations du modèle précédent, pour pallier le problème de surdétermination qui survient lorsque la dimension du processus observé est supérieure à deux. Les techniques d'estimation sont toujours fondées sur la variation quadratique réalisée, et nous proposons d'autres outils afin de continuer à estimer les composantes de la volatilité avec la vitesse optimale en présence des termes d'erreur. Des tests numériques permettent de mettre en évidence la présence de telles erreurs dans nos données.

Enfin, dans la dernière partie nous résolvons le problème d'un producteur qui intervient sur le marché infrajournalier de l'électricité afin de compenser les coûts liés aux rendements aléatoires de ses unités de production. Par ses actions, il exerce un impact sur le marché. Les prix et son anticipation de la demande de ses consommateurs sont modélisés par une diffusion à sauts. Les outils du contrôle optimal stochastique permettent de déterminer sa stratégie dans un problème approché. Nous donnons des conditions pour que cette stratégie soit très proche de l'optimalité dans le problème de départ, et l'illustrons numériquement.

Mots-clés

Efficacité asymptotique ; erreurs de modèle ; estimation non paramétrique par noyaux ; exécution optimale ; impact de marché ; prix de l'électricité ; programmation dynamique ; statistique des diffusions ; volatilité intégrée.

Abstract

In this thesis, we study mathematical models for the representation of prices on the electricity markets, from the viewpoints of statistics of random processes and optimal stochastic control.

In a first part, we perform estimation of the components of the volatility coefficient of a multidimensional diffusion process, which represents the evolution of prices in the electricity forward market. It is driven by two Brownian motions. We aim at achieving estimation efficiently in terms of convergence rate and, concerning the parametric part of those components, in terms of limit law. To do so, we must extend the usual notion of efficiency in the Cramér-Rao sense. Our estimation methods are based on realized quadratic variation of the observed process.

In a second part, we add model error terms to the previous model, in order to care for some kind of degeneration occurring in it as soon as the dimension of the observed process is greater than two. Our estimation methods are still based on realized quadratic variation, and we give other tools in order to keep on estimating the volatility components with the optimal rate when error terms are present. Then, numerical tests provide us with some evidence that such errors are present in the data.

Finally, we solve the problem of a producer, which trades on the electricity intraday market in order to cope with the uncertainties on the outputs of his production units. We assume that there is market impact, so that the producer influences prices as he trades. The price and the forecast of the consumers' demand are modelled by jump diffusions. We use the tools of optimal stochastic control to determine the strategy of the producer in an approximate problem. We give conditions so that this strategy is close to optimality in the original problem, as well as numerical illustrations of that strategy.

Keywords

Asymptotic efficiency; dynamic programming; electricity prices; estimation for diffusions; integrated volatility; market impact; model errors; nonparametric kernel estimation; optimal execution.

Cette thèse de doctorat a été préparée au sein du Laboratoire de Probabilités et Modèles Aléatoires (LPMA) et de l'unité de formation et de recherche de mathématiques de l'université Paris Diderot, sise à l'adresse suivante : Case courrier 7012 / Avenue de France / 75105 PARIS CEDEX 13.

Table des matières

Chapitre 1

Introduction

1.1 Introduction

Dans ces travaux, nous étudions des questions de statistique des processus et de contrôle optimal stochastique, dans un cadre de modélisation des prix sur les marchés de l'électricité. Nous nous intéressons aux marchés de gros, sur lesquels interviennent des producteurs et des fournisseurs d'électricité, ainsi que des négociants intermédiaires ; les clients finaux, particuliers comme entreprises, n'y ont pas accès. Ces marchés se distinguent par l'horizon temporel des contrats qui s'y échangent et, du fait que l'électricité ne se stocke pas, un contrat donnant lieu à la livraison d'électricité comportera toujours la précision de la période de livraison sous-jacente.

Dans une première partie, nous nous placerons sur le marché à terme de l'électricité, qui est un marché financier classique, et où nous disposons de données historiques de prix de contrats à terme. À partir de celles-ci, nous réaliserons de l'estimation statistique pour un processus de diffusion multidimensionnel à volatilité stochastique représentant la dynamique de ces prix. La spécification de la structure de volatilité est guidée par l'observation empirique que sur ce marché, les prix des contrats deviennent plus volatils à l'approche de leur maturité. La quantification de l'accroissement de la volatilité sera réalisée à travers l'estimation d'un paramètre réel apparaissant dans sa structure. Nous réaliserons cette estimation *efficacement*, dans un sens voisin de celui de Cramér-Rao qu'il nous faudra préciser, en nous appuyant sur la théorie classique de l'estimation en statistique semi-paramétrique telle qu'elle est présentée, par exemple, dans le 25e chapitre du livre de Van der Vaart [73]. Nous nous attacherons également à estimer non paramétriquement les trajectoires des autres composantes de la volatilité, qui sont des processus stochastiques sur lesquels nous ferons des hypothèses de régularité. Les performances des estimateurs sont ensuite évaluées sur des jeux de données simulées et réelles.

Dans une deuxième partie, nous travaillerons toujours sur le modèle de diffusion pour les prix de l'électricité sur le marché à terme, mais les observations incorporeront des *erreurs de modèle*. Leur introduction est motivée par une dégénérescence du modèle, observée dans la partie précédente, qui se produit lorsque la dimension du processus observé est supérieure au nombre de mouvements browniens conduisant sa dynamique. Notre approche comporte des similitudes avec l'estimation en présence de bruit de microstructure telle que décrite, par exemple, dans l'introduction de Zhang *et al.* [77]. Dans notre modélisation, nous nous inspirerons des techniques de traitement du bruit de microstructure bien que le cadre méthodologique soit différent ; les termes d'erreur seront asymptotiquement petits, ce qui permettra d'utiliser les estimateurs de la première partie pour estimer les composantes du processus de diffusion bruité. Nous mettrons en valeur la manière dont les propriétés des estimateurs sont, toutefois, affectées par l'ajout des termes d'erreur. Ceux-ci influent sur les vitesses de convergence des estimateurs, mais aussi sur leurs lois limites. Nous proposerons alors de nouveaux estimateurs, qui permettent d'estimer les composantes paramétrique et non paramétrique avec les vitesses optimales de leurs paradigmes respectifs. Enfin, nous confronterons nos procédures d'estimation à des tests sur données simulées pour comprendre comment l'enrichissement du modèle par l'ajout de termes d'erreur peut permettre de comprendre les comportements des estimateurs sur les données historiques du marché.

Dans la dernière partie de ces travaux, nous appliquerons des techniques de contrôle optimal stochastique des diffusions incorporant des sauts, afin de résoudre le problème d'un acteur sur le marché infrajournalier de l'électricité. Cet acteur, doté de moyens de production au rendement aléatoire et cherchant à anticiper la demande future de consommateurs dont il a la charge, se place sur le marché infrajournalier pour acheter ou vendre, en temps continu, des contrats donnant lieu à la livraison d'électricité afin d'atteindre sa cible de consommation. Être éloigné de cette cible donne lieu à des pénalités. De plus, nous considérerons que l'action de l'acteur a une influence sur le marché, qui n'est pas parfaitement liquide, par le biais du modèle d'impact de marché d'Almgren et Chriss [12]. Nous déterminerons la stratégie d'achat/vente de l'acteur grâce au principe de la programmation dynamique, avec le souci d'obtenir des formules explicites pour en retirer des interprétations économiques. Cela ne sera pas possible dans le problème général, aussi nous obtiendrons des expressions analytiques dans un problème approché, et nous montrerons leur proximité vis-à-vis des solutions du problème général. Des expériences numériques illustreront les stratégies que nous dériverons.

Dans cette introduction, nous présentons la problématique et les principaux résultats de chacune des trois parties, puis nous décrivons les marchés sur lesquels nous avons travaillé et les données utilisées.

1.2 Première partie
Estimation efficace dans un modèle à deux facteurs

Nous considérons dans cette partie un processus multidimensionnel X, défini sur l'espace filtré $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, de composantes $X^1, ..., X^d$ vérifiant, pour $j = 1, ..., d$ et $t \geq 0$,

$$X_t^j = X_0^j + \int_0^t b_s^j ds + \int_0^t e^{-\vartheta(T_j - s)}\sigma_s dB_s + \int_0^t \overline{\sigma}_s d\overline{B}_s, \tag{1.1}$$

où $X_0^j \in \mathbb{R}$ est une condition initiale, $B = (B_t)_{t \geq 0}$ et $\overline{B} = (\overline{B}_t)_{t \geq 0}$ sont deux mouvements browniens indépendants, ϑ et T_j sont deux réels strictement positifs, et $\sigma = (\sigma_t)_{t \geq 0}$, $\overline{\sigma} = (\overline{\sigma}_t)_{t \geq 0}$, $b^j = (b_t^j)_{t \geq 0}$ sont des processus càdlàg et adaptés. De plus, les T_j vérifient

$$T \leq T_1 < \cdots < T_d,$$

pour un certain $T > 0$, et nous disposons d'observations discrètes $X_{iT/n}^1, ..., X_{iT/n}^d$ synchrones, pour $i = 0, ..., n$. Le schéma d'observation est donc régulier, et dans notre asymptotique le pas d'observation $\Delta_n = T/n$ est petit devant T, c'est-à-dire que T/Δ_n tend vers $+\infty$.

Ce modèle a vocation à être appliqué aux logarithmes de prix de contrats à terme sur l'électricité, présentés dans la section 1.5 : l'entier d représentera le nombre de contrats dont on observe simultanément les prix. Pour nous, d variera entre 2 et 6. Les dates 0 et T correspondront aux premier et dernier instants auxquels ces d prix sont observables. La longueur de la période $[0, T]$ sera appelée à varier entre un et cinq mois selon la valeur de d. Notre cadre asymptotique peut donc être considéré de la manière suivante : Δ_n est de l'ordre d'une journée, que nous considérerons petite devant T, qui est de l'ordre de quelques mois. Les nombres $T_1, ..., T_d$ sont des dates, exprimées en jours, correspondant aux premiers jours de mois consécutifs pour lesquels l'un des d contrats, donnant lieu à la livraison d'un mégawatt-heure répartie sur la durée du mois, est coté sur le marché à terme sur la période $[0, T]$. La dynamique du logarithme du prix du contrat $j = 1, ..., d$ est donnée par (1.1).

Notre problématique est l'estimation des composantes apparaissant dans les coefficients de volatilité. Précisément, nous cherchons à estimer le paramètre réel ϑ ainsi que les trajectoires des processus σ et $\overline{\sigma}$. L'enjeu essentiel est l'estimation efficace de ϑ, c'est-à-dire que nous cherchons à l'estimer avec la meilleure vitesse et le meilleur théorème limite possibles. Nos méthodes d'estimation s'appuieront sur les techniques de traitement des données haute-fréquence. Nous cherchons également des conditions pour estimer σ et $\overline{\sigma}$ avec la vitesse usuelle.

Pour un processus de diffusion très général

$$dX_t = b_t dt + \sigma_t dW_t$$

observé sur un intervalle $[0, T]$ fixé, il est connu (voir, par exemple, la remarque à ce propos dans le chapitre d'ouvrage de Jacod [49], p. 198) que la dérive b ne peut être estimée. Il est uniquement possible d'estimer des composantes de la volatilité σ.

La question de l'estimation du coefficient de volatilité a fait l'objet de nombreux travaux ; dans le cas où σ est une fonction connue de t, de X_t et d'un paramètre m-dimensionnel ϑ, Genon-Catalot et Jacod [33] dérivent des estimateurs de ϑ. Genon-Catalot *et al.*. [34] ont étudié l'estimation non paramétrique de la volatilité dans le cas où X est unidimensionnel, b est une fonction déterministe de t et de X_t, et σ est une fonction déterministe de t. L'estimation est réalisée à l'aide de méthodes d'ondelettes. Par la suite, Hoffmann [43] a estimé non paramétriquement la volatilité d'un processus de diffusion unidimensionnel, sous une spécification déterministe, à la vitesse minimax sous l'hypothèse que la volatilité appartient à une certaine classe de Besov. Des résultats étendus à des classes de régularité plus larges ont été obtenus par Hoffmann [44].

Dans le même temps, le besoin d'adopter des modèles plus généraux a conduit à considérer que le coefficient de volatilité peut être lui-même un processus stochastique. Dès lors, il n'est plus possible de parler de son estimation au sens usuel, car σ est lui-même aléatoire ; par exemple, étant donné des observations discrètes d'un processus de diffusion, il ne s'agit pas d'estimer non paramétriquement $t \rightsquigarrow \sigma_t$, mais plutôt la trajectoire $t \rightsquigarrow \sigma_t(\omega)$, qui est différente d'un jeu de données à l'autre. Pour établir des résultats d'estimation non paramétrique, il est nécessaire de supposer des conditions de régularité sur les trajectoires du processus σ. On se réfère ici à la discussion dans l'introduction de Jacod [49] : si l'on souhaite estimer la volatilité intégrée $\int_0^t \sigma_s^2 ds$ pour $t \in [0, T]$, qui est un nombre (aléatoire aussi), ces hypothèses de régularité sur σ ne sont pas nécessaires et l'on peut essayer, comme en statistique paramétrique classique, de réaliser l'estimation avec la meilleure vitesse de convergence et la meilleure loi limite possible (dans un sens à préciser). Par ailleurs, connaître la volatilité intégrée de 0 à t, pour tout t, permet d'avoir une idée de σ^2, et donc de σ.

De fait, estimer la volatilité intégrée est devenu un problème emblématique. Les statisticiens ont dû se munir de nouveaux outils pour pouvoir, par exemple, énoncer des théorèmes limites où la limite dépend de la trajectoire du processus estimé ; fondés sur le concept de convergence stable en loi, introduit par Rényi [68] dès 1963, et sur la théorie des semi-martingales, des lois des grands nombres et des théorèmes limites ont été développés pour estimer la volatilité intégrée, mais aussi, plus largement, des intégrales de fonctions de la volatilité y compris pour des processus multidimensionnels. Ces évolutions ont été conduites par des travaux probabilistes comme celui de Jacod [48], qui établit des conditions pour qu'une famille de semi-martingales converge stablement en loi vers un processus limite. Une référence concernant les théorèmes limites est le livre de Jacod et Shiryaev [54]. D'un point de vue statistique, les deux chapitres d'ouvrage de Mykland et Zhang [62] et Jacod [49] présentent les problématiques les plus importantes autour de l'estimation des diffusions sur un intervalle de temps fixé, ainsi que des résultats d'estimation centraux énoncés dans un cadre très général.

La question de l'optimalité de procédures d'estimation de la volatilité intégrée s'est posée ; attendu qu'elle peut être estimée à la vitesse $\sqrt{n}$, la qualité des estimateurs que l'on en propose se mesure à la loi limite que l'on peut obtenir dans un théorème de la limite centrale. Clément *et al.* [24] ont traité la question de l'estimation de fonctionnelles de la volatilité ; dans le modèle de diffusion qu'ils introduisent, ils prouvent une extension du théorème de convolution de Hájek et définissent ainsi une notion d'efficacité. Jacod et Rosenbaum [53] estiment également des fonctionnelles de la volatilité cherchent notamment des estimateurs efficaces dans le modèle de Clément *et al.*.

Ces dernières années, motivés notamment par des applications financières, les chercheurs ont orienté leurs efforts vers l'estimation dans des processus comportant une partie diffusive et des sauts, à partir d'observations bruitées. Nous reportons une discussion approfondie sur ce point à la section suivante.

À propos du choix du modèle statistique sur lequel nous allons travailler, Hinz [42] a établi un lien méthodologique entre les produits de taux d'intérêt et les contrats à terme sur l'électricité, justifiant d'appliquer à ce dernier champ des modèles issus de travaux sur les taux. Notamment, le modèle de Heath-Jarrow-Morton (HJM), introduit dans Heath *et al.* [40] a eu du succès dans cette voie, comme en témoigne l'importante étude de modélisation réalisée par Benth et Koekebakker [16] et l'utilisation de ce modèle dans des travaux comme ceux de Kiesel *et al.* [57]. C'est d'ailleurs de ces derniers travaux que le modèle de ce chapitre provient. Il apparaît aussi dans le chapitre 11 du livre de Musiela et Rutkowski [61], pour représenter la dynamique des taux à terme.

Dans un contexte de taux d'intérêts, Bhar *et al.* [19] ont réalisé de l'estimation dans un processus de HJM avec une spécification de la volatilité entièrement déterministe. Une spécification moins contrainte est celle de Jeffrey *et al.* [55], qui modélise le rendement de zéro-coupons par un processus de HJM dont la volatilité est une fonction non spécifiée du taux à court terme. C'est cette fonction que les auteurs estiment, dans un paradigme non paramétrique.

Dans notre modèle de diffusion particulier, nous nous posons les questions suivantes :

Question 1 Peut-on écrire un estimateur de ϑ efficace, dans un sens voisin de celui des travaux de Clément *et al.* [24] et Jacod et Rosenbaum [53] ?

Question 2 Peut-on estimer non paramétriquement les trajectoires de σ et $\overline{\sigma}$, et est-il possible de définir et d'effectuer une estimation à la vitesse optimale, au sens minimax, de ces trajectoires ?

Question 3 Quelle est la configuration, dictée par le nombre de processus observés comme exposé dans la section 1.5, dans laquelle les conditions d'estimation de ϑ sont les plus favorables ?

La question 1 revient ici à se demander si pour un sous-modèle avec des processus de volatilité déterministes, il est possible d'estimer ϑ efficacement au sens usuel de Cramér-Rao. Nous qualifierons d'efficace une procédure d'estimation ayant cette propriété. Notre problème d'estimation efficace est cependant différent de ceux que nous avons cités ci-dessus, car nous n'estimons pas une quantité liée à la volatilité, qui est un processus non spécifié ; nous nous posons la question de l'estimation du paramètre réel ϑ. La question 2 se pose naturellement, à la suite des travaux d'Hoffmann [44, 45] où les classes de Besov étaient les outils naturels pour définir une théorie minimax de l'estimation d'une fonction de volatilité déterministe. Enfin, la question 3 nécessitera une confrontation à des jeux de données simulées et réelles, en regard des résultats théoriques de convergence d'estimateurs de ϑ.

Nous commençons par distinguer les problèmes d'estimation statistique suivant la dimension d du processus X : si $d = 1$, le triplet $(\vartheta, \sigma, \overline{\sigma})$ n'est pas identifiable. Le problème statistique le plus régulier est obtenu quand $d = 2$: nous pouvons alors utiliser des techniques fondées sur l'approximation de la variation quadratique pour proposer un estimateur de ϑ ; en effet, comme

$$d(X_t^2 - X_t^1) = (b_t^2 - b_t^1)dt + \left(e^{-\vartheta T_2} - e^{-\vartheta T_1}\right)e^{\vartheta t}\sigma_t dB_t,$$

le résultat classique d'estimation de la volatilité intégrée donne que

$$\Psi^n_{T_1,T_2} = \frac{\sum_{i=1}^n (\Delta_i^n X^2 - \Delta_i^n X^1)^2}{\sum_{i=1}^n \left((\Delta_i^n X^2)^2 - (\Delta_i^n X^1)^2\right)} \to \frac{\left(e^{-\vartheta T_2} - e^{-\vartheta T_1}\right)^2}{e^{-2\vartheta T_2} - e^{-2\vartheta T_1}} = \psi_{T_1,T_2}(\vartheta)$$

en probabilité, quand $n \to \infty$, où $\Delta_i^n X^j = X^j_{i\Delta_n} - X^j_{(i-1)\Delta_n}$, $j = 1, 2$. On définit l'estimateur $\hat{\vartheta}_{2,n}$ comme étant égal à $\psi^{-1}_{T_1,T_2}\big(\Psi^n_{T_1,T_2}\big)$ quand cela est possible. Les résultats de Jacod [49] permettent d'établir ce qui suit dans la première partie du théorème 2.1 :

Résultat 1. *Nous avons la convergence*

$$\Delta_n^{-1/2}(\hat{\vartheta}_{2,n} - \vartheta) \to \mathcal{N}\big(0, V_\vartheta(\sigma, \overline{\sigma})\big)$$

stablement en loi quand $n \to \infty$, où $\mathcal{N}\big(0, V_\vartheta(\sigma, \overline{\sigma})\big)$ est une variable aléatoire qui, conditionnellement à $\mathcal{F}$, est gaussienne, centrée, de variance

$$V_\vartheta(\sigma, \overline{\sigma}) = \frac{1}{(T_2 - T_1)^2}(e^{\vartheta T_2} - e^{\vartheta T_1})^2 \frac{\int_0^T e^{2\vartheta t}\sigma_t^2 \overline{\sigma}_t^2 dt}{\big(\int_0^T e^{2\vartheta t}\sigma_t^2 dt\big)^2}.$$

Cela permet d'obtenir que l'estimation de ϑ peut être faite à la vitesse usuelle $\Delta_n^{-1/2}$ du paradigme paramétrique, sous la seule hypothèse que les processus b, σ et $\overline{\sigma}$ soient càdlàg, adaptés, et que σ et $\overline{\sigma}$ soient strictement positifs presque-sûrement.

Le cas $d \geq 3$ présente une forme de dégénérescence, du fait que la dimension du processus X est strictement supérieure au nombre de mouvements browniens conduisant sa dynamique : dans le cas particulier où les processus de dérive b^1, b^2 et b^3 sont égaux, nous avons l'égalité

$$\frac{\Delta_1^n X^3 - \Delta_1^n X^2}{\Delta_1^n X^2 - \Delta_1^n X^1} = \frac{e^{-\vartheta T_3} - e^{-\vartheta T_2}}{e^{-\vartheta T_2} - e^{-\vartheta T_1}}$$

où le membre de droite définit une fonction inversible de ϑ. Il s'ensuit qu'observer un seul incrément du processus tridimensionnel X suffit à obtenir ϑ sans erreur. Si les processus de dérive ne sont pas tous égaux, le problème demeure dégénéré car il reste possible de supprimer l'aléa lié aux mouvements browniens, et dès lors, ce sont les termes de dérive qui vont dicter la vitesse de convergence des estimateurs de ϑ.

On introduit à ce moment-là un estimateur qui exploite cette singularité : on a la convergence

$$\Psi^n_{T_1,T_2,T_3} = \frac{\sum_{i=1}^n (\Delta_i^n X^3 - \Delta_i^n X^2)^2}{\sum_{i=1}^n (\Delta_i^n X^2 - \Delta_i^n X^1)^2} \to \Big(\frac{e^{-\vartheta T_3} - e^{-\vartheta T_2}}{e^{-\vartheta T_2} - e^{-\vartheta T_1}}\Big)^2 = \psi_{T_1,T_2,T_3}(\vartheta)$$

en probabilité, et donc, en définissant $\hat{\vartheta}_{3,n}$ comme $\psi^{-1}_{T_1,T_2,T_3}\big(\Psi^n_{T_1,T_2,T_3}\big)$ lorsque cela est possible, on construit un estimateur convergent de ϑ. Dans la deuxième partie du théorème 2.1, sous l'hypothèse que pour un certain $s > 1/2$, l'on a la majoration

$$\sup_{t \in [0,T]} t^{-s} \omega(b^j)_t < \infty \text{ pour tout } j = 1, 2, 3, \tag{1.2}$$

où

$$\omega(X)_t = \sup_{|h| \leq t} \Big(\int_0^T \mathbb{E}\big[(X_{s+h} - X_s)^2\big] ds\Big)^{1/2}, \tag{1.3}$$

on obtient le résultat suivant.

Résultat 2. *La suite* $\left(\Delta_n^{-1}(\hat{\vartheta}_{3,n} - \vartheta)\right)_{n\geq 1}$ *est tendue en probabilité.*

Dans la variable aléatoire limite apparaissent ϑ, σ et les processus de dérive b^1, b^2, b^3. Ce résultat est établi en utilisant le lemme technique 2.4.1, reposant sur des résultats d'approximation stochastique. Pour des processus déterministes, l'hypothèse technique (1.2) reviendrait à dire que b^1, b^2 et b^3 appartiennent à des espaces de Besov $\mathcal{B}^s_{2,\infty}([0,T])$, avec $s > 1/2$. Ici, nous avons donc une forme de régularité de Besov formulée en espérance, par le biais du module de continuité (1.3). Nous nous plaçons ensuite dans un cadre d'approximation dans la base de Haar pour établir le lemme.

À ce stade, nous ne sommes pas parvenus à répondre à la question 3 posée plus haut ; dans la section 2.2.4, nous donnons un estimateur $\hat{\vartheta}_{d,n}$ de ϑ dans le cas général $d \geq 3$, et la proposition 2.1 indique que $\Delta_n^{-1}(\hat{\vartheta}_{d,n} - \vartheta)$ converge en probabilité vers une certaine variable aléatoire. Un moyen simple de fournir une réponse à la question 3 aurait été de donner des conditions pour comparer les limites en probabilité de $\Delta_n^{-1}(\hat{\vartheta}_{3,n} - \vartheta), ..., \Delta_n^{-1}(\hat{\vartheta}_{d,n} - \vartheta)$ en indiquant laquelle est la plus proche de 0. Notons cependant que l'estimateur $\hat{\vartheta}_{2,n}$ converge vers ϑ à la vitesse $\Delta_n^{-1/2}$ alors que $\hat{\vartheta}_{3,n}, ..., \hat{\vartheta}_{d,n}$ convergent à la vitesse plus élevée Δ_n^{-1}. Ces derniers sont asymptotiquement biaisés, mais ils sont plus rapides. À distance finie, spécifiquement quand il y a peu de données, ces comparaisons ne sont plus si évidentes.

Nous établissons ensuite un résultat d'estimation non paramétrique, en travaillant dans le cadre régulier $d = 2$. En utilisant les résultats classiques d'estimation de la volatilité intégrée, nous avons la convergence en probabilité

$$\sum_{i=1}^{n} g((i-1)\Delta_n)(\Delta_i^n X^j)^2 \to \int_0^T g(s)\left(e^{-2\vartheta(T_j - s)}\sigma_s^2 + \overline{\sigma}_s^2\right)ds,$$

pour $j = 1, 2$, dès lors que g est suffisamment régulière. En choisissant une suite de fonctions $(g_n)_n$ se rapprochant de plus en plus de la masse de Dirac au point $t \in [0,T]$, l'on devrait obtenir que

$$\sum_{i=1}^{n} g_n((i-1)\Delta_n)(\Delta_i^n X^j)^2 \to e^{-2\vartheta(T_j - t)}\sigma_t^2 + \overline{\sigma}_t^2,$$

de sorte que l'on peut, asymptotiquement, estimer les carrés des volatilités équivalentes des processus X^1 et X^2 à l'instant t. Par ailleurs, nous avons la relation simple

$$\begin{pmatrix} \sigma_t^2 \\ \overline{\sigma}_t^2 \end{pmatrix} = \mathcal{M}(\vartheta)_t \begin{pmatrix} e^{-2\vartheta(T_1 - t)}\sigma_t^2 + \overline{\sigma}_t^2 \\ e^{-2\vartheta(T_2 - t)}\sigma_t^2 + \overline{\sigma}_t^2 \end{pmatrix},$$

où

$$\mathcal{M}(\vartheta)_t = \frac{1}{e^{-2\vartheta(T_1 - t)} - e^{-2\vartheta(T_2 - t)}} \begin{pmatrix} 1 & -1 \\ -e^{-2\vartheta(T_2 - t)} & e^{-2\vartheta(T_1 - t)} \end{pmatrix}.$$

On propose donc les estimateurs

$$\begin{pmatrix} \widehat{\sigma}_{n,t}^2 \\ \widehat{\overline{\sigma}}_{n,t}^2 \end{pmatrix} = \mathcal{M}(\max\{\hat{\vartheta}_{2,n}, \varpi_n\})_t \sum_{i=1}^{n} K_{h_n}(t - (i-1)\Delta_n) \begin{pmatrix} (\Delta_i^n X^1)^2 \\ (\Delta_i^n X^2)^2 \end{pmatrix},$$

où $(\varpi_n)_n$ est une suite de réels strictement positifs décroissant vers 0, car $\mathcal{M}(\hat{\vartheta}_{2,n})_t$ n'est pas définie si $\hat{\vartheta}_{2,n}$ prend la valeur 0. On prend par ailleurs $K_{h_n}(t) = h_n^{-1}K(t/h_n)$, où $K(t) = \mathbf{1}_{(0,1]}(t)$. Ce noyau est dit *causal* car son support est inclus dans $(0, \infty)$; les estimateurs non paramétriques des processus de volatilité aux instants $i\Delta_n$ sont donc $\mathcal{F}_{i\Delta_n}$-mesurables, ce qui est une propriété technique importante sur laquelle nous reviendrons.

Sous l'hypothèse que, pour un certain réel positif c et pour un réel $\alpha \geq 1/2$, l'on a

$$\mathbb{E}\big[|\sigma_t^2 - \sigma_s^2|^2\big] + \mathbb{E}\big[|\overline{\sigma}_t^2 - \overline{\sigma}_s^2|^2\big] \leq c|t-s|^{2\alpha}, \tag{1.4}$$

ce qui s'apparente à une régularité de Hölder formulée en espérance, le théorème 2.2 établit alors le résultat suivant.

Résultat 3. *La suite*

$$\Big(\Delta_n^{-\alpha/(2\alpha+1)}\Big[\big|\widehat{\sigma}_{n,t}^2 - \sigma_t^2\big| + \big|\widehat{\overline{\sigma}}_{n,t}^2 - \overline{\sigma}_t^2\big|\Big]\Big)_{n\geq 1}$$

est tendue en probabilité, uniformément pour t dans un compact $\mathcal{D}$ inclus dans $(0, T]$.

En ce sens, nous atteignons la vitesse minimax d'estimation non paramétrique, ce qui répond à la question 2.

Pour établir le résultat d'estimation efficace, nous travaillons là encore dans le cas $d = 2$. Nous nous plaçons temporairement dans un sous-modèle de notre modèle initial, où σ et $\overline{\sigma}$ sont des fonctions déterministes et strictement positives. L'on utilise alors la théorie de l'estimation semi-paramétrique (voir le 25^e^ chapitre de Van der Vaart [73]) pour dériver la fonction de score efficace dans ce sous-modèle à partir de projections sur des espaces tangents. La fonction de score efficace associée à l'observation de $(\Delta_i^n X^1, \Delta_i^n X^2)$, pour $i = 1, ..., n$, est donnée par

$$\widetilde{\ell}_{\overline{\sigma}}(\vartheta)^i = \frac{(\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta(T_2-T_1)}\Delta_i^n X^1)e^{-\vartheta(T_2-T_1)}(T_2-T_1)}{(1-e^{-\vartheta(T_2-T_1)})^3\int_{(i-1)\Delta_n}^{i\Delta_n}\overline{\sigma}_t^2 dt}. \tag{1.5}$$

Nous obtenons alors que dans le sous-modèle avec des volatilités déterministes, la borne de Cramér-Rao pour estimer ϑ est donnée par

$$V_\vartheta^{\mathrm{opt}}(\sigma, \overline{\sigma}) = \frac{1}{(T_2-T_1)^2}(e^{\vartheta T_2} - e^{\vartheta T_1})^2\Big(\int_0^T \frac{e^{2\vartheta t}\sigma_t^2}{\overline{\sigma}_t^2}dt\Big)^{-1}.$$

Ce résultat est donné dans le théorème 2.3. Toujours dans ce sous-modèle, une procédure d'estimation serait dite efficace si elle atteignait cette borne. Nous étendons ce concept au modèle dans sa globalité, en disant que l'estimation est réalisée efficacement si elle atteint la borne de Cramér-Rao au sens classique quand les processus de volatilité sont déterministes.

Pour répondre à la question 1, nous exhibons un estimateur à un pas $\tilde{\vartheta}_{2,n}$, défini par

$$\tilde{\vartheta}_{2,n} = \hat{\vartheta}_{2,n} + \frac{\sum_{(i-1)\Delta_n\in[h_n,T]}\widetilde{\ell}\big(\hat{\vartheta}_{2,n}, \widehat{\overline{\sigma}}_n^2\big)^i}{\sum_{(i-1)\Delta_n\in[h_n,T]}\big(\widetilde{\ell}\big(\hat{\vartheta}_{2,n}, \widehat{\overline{\sigma}}_n^2\big)^i\big)^2},$$

où

$$\widetilde{\ell}\big(\hat{\vartheta}_{2,n},\widehat{\overline{\sigma}}_n^2\big)^i = \frac{(\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\hat{\vartheta}_{2,n}(T_2-T_1)}\Delta_i^n X^1)e^{-\hat{\vartheta}_{2,n}(T_2-T_1)}(T_2-T_1)}{(1-e^{-\hat{\vartheta}_{2,n}(T_2-T_1)})^3\Delta_n\widehat{\overline{\sigma}}^2_{n,(i-1)\Delta_n}}$$

est obtenu en remplaçant, dans l'expression (1.5), le paramètre ϑ par l'estimateur $\hat{\vartheta}_{2,n}$ et l'intégrale inconnue $\int_{(i-1)\Delta_n}^{i\Delta_n}\overline{\sigma}_t^2 dt$ par l'expression $\Delta_n\widehat{\overline{\sigma}}^2_{n,(i-1)\Delta_n}$ fournie par le résultat d'estimation non paramétrique afin que l'estimateur puisse effectivement être calculé.

La principale difficulté consiste à prouver que la différence entre, d'une part, la somme des scores efficaces $\widetilde{\ell}_{\overline{\sigma}}(\vartheta)^i$ et d'autre part, la somme des quantités analogues $\widetilde{\ell}\big(\hat{\vartheta}_{2,n},\widehat{\overline{\sigma}}_n^2\big)^i$ dans lesquelles les estimateurs non paramétriques ont été insérés, est suffisamment petite quand $n\to\infty$. Nous sommes ainsi amenés à prouver le théorème 2.4, dans lequel on trouvera le résultat suivant.

Résultat 4. *Nous avons la convergence*

$$\Delta_n^{-1/2}\big(\tilde{\vartheta}_{2,n}-\vartheta\big)\to\mathcal{N}\big(0,V_\vartheta^{\mathrm{opt}}(\sigma,\overline{\sigma})\big)$$

stablement en loi quand $n\to\infty$. Conditionnellement à $\mathcal{F}$, la loi limite est gaussienne, centrée, de variance $V_\vartheta^{\mathrm{opt}}(\sigma,\overline{\sigma})$.

L'estimation efficace est, en ce sens, réalisée. Pour établir ce résultat, nous devons renforcer l'hypothèse (1.4) en requérant que $\alpha > 1/2$. Nous avons également besoin d'une hypothèse assurant que σ et $\overline{\sigma}$ sont presque-sûrement bornés inférieurement par une constante $\tilde{c}>0$. Enfin, le choix d'un noyau causal pour l'estimation non paramétrique permet d'obtenir des estimateurs de $\overline{\sigma}^2_{i\Delta_n}$ qui sont $\mathcal{F}_{i\Delta_n}$-mesurables, et cette propriété est exploitée de manière cruciale pour prouver les résultats d'approximation stochastique.

Nous menons ensuite des expériences numériques pour tester le comportement de nos estimateurs sur des données simulées puis réelles. Les conditions dans lesquelles nous simulons des jeux de données sont dictées par les schémas d'observation des prix sur le marché à terme de l'électricité, que nous décrirons plus bas.

Sur des données simulées, nous mettons en avant la capacité des estimateurs $\hat{\vartheta}_{2,n},\ldots,\hat{\vartheta}_{d,n}$ et $\tilde{\vartheta}_{2,n}$ à discerner un paramètre ϑ significativement non nul dans les conditions d'observation des données réelles. Il apparaît aussi que nous pouvons exhiber des configurations dans lesquelles considérer $d>3$ processus de prix pour l'estimation conduit à un biais plus limité qu'avec 3 processus seulement, et d'autres dans lesquelles l'effet inverse se produit ; il n'y a donc à ce stade pas de réponse claire à la question 3.

Par ailleurs, nous testons l'estimateur non paramétrique en réalisant 10 000 simulations conduisant à autant de courbes d'estimateurs non paramétriques. Nous traçons ensuite la courbe reliant les moyennes de tous les estimateurs en un point de la grille, ainsi que les courbes des quantiles à $2,5\%$ et à $97,5\%$. Nous obtenons la figure 1.1 dans un cas où les volatilités sont déterministes et nous autorisent donc à représenter la vraie volatilité équivalente de X^1. Outre les écarts constatés en début de courbe du fait du choix du noyau

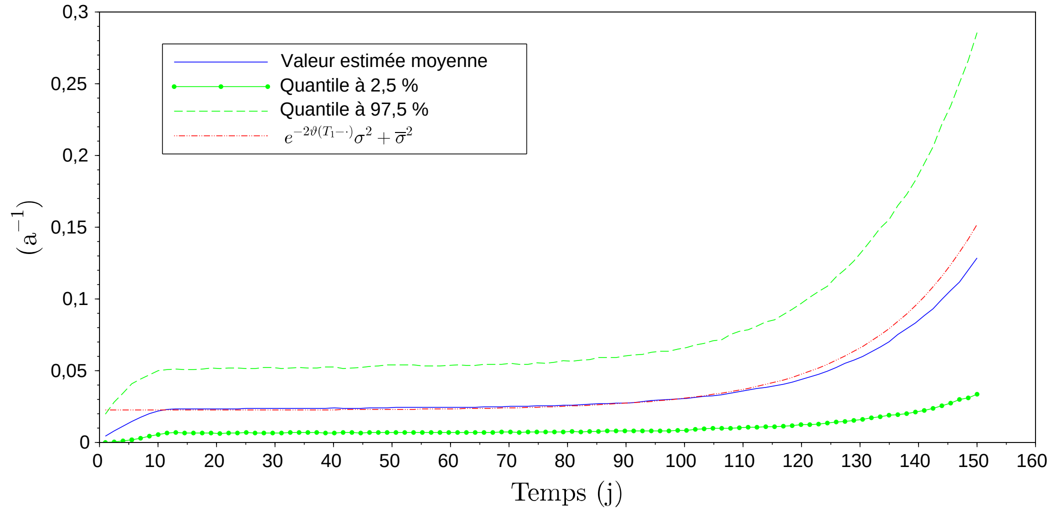

FIGURE 1.1 – Moyenne et quantiles pour l'estimation du carré de la volatilité équivalente du processus X^1

causal et de la fenêtre d'estimation, la qualité de l'approximation du carré de la volatilité équivalente par les courbes $e^{-2\hat{\vartheta}_{2,n}(T_1-t)}\widehat{\sigma}^2_{n,t}+\widehat{\overline{\sigma}}^2_{n,t}$ est satisfaisante.

Enfin, des expériences sur données réelles montrent que les estimateurs $\hat{\vartheta}_{2,n}$ et $\tilde{\vartheta}_{2,n}$ fournissent des valeurs bien plus élevées que celles de $\hat{\vartheta}_{3,n},...,\hat{\vartheta}_{d,n}$. Cela est en désaccord avec les résultats sur données simulées, et cette confrontation nous mène à considérer la notion d'*erreur de modèle*, que nous cherchons à quantifier dans la partie suivante.

1.3 Deuxième partie
Estimation en présence d'erreurs de modèle

Nous considérons le processus multidimensionnel X de la partie précédente. Nous ne l'observons cependant plus de manière directe ; les observations, discrètes et synchrones, sont $Y^1_{i\Delta_n},...,Y^d_{i\Delta_n}$, pour $i=0,...,n$. Elles vérifient, pour $j=1,...,d$,

$$Y^j_{i\Delta_n}=X^j_{i\Delta_n}+\kappa^n_j\epsilon^j_i,$$

où les ϵ^j_i sont des variables aléatoires i.i.d., centrées, ne dépendant pas du processus X. Pour $j=1,...,d$, $(\kappa^n_j)_n$ est une suite de réels positifs décroissant vers 0, et vérifiant de plus $n^\beta\kappa^n_j\to\iota_j$, pour un réel $\beta\geq 1/2$ et un certain $\iota_j>0$. L'asymptotique est prise quand $\Delta_n=T/n$ tend vers 0. Le contexte est donc celui de bruits additifs qui, asymptotiquement, ne sont pas plus grands que le processus X, dont les incréments sont d'ordre $\Delta_n^{1/2}$.

La motivation pour l'introduction d'erreurs de modèle vient du problème de *surdétermination* constaté dans la partie précédente ; dès que $d \geq 3$, les d composantes du processus X sont conduites par deux aléas browniens seulement. Cette difficulté est liée aux modèles de type HJM, et a également été notée dans le cadre de la modélisation de la courbe des taux, par exemple par Jeffrey *et al.* [55]. L'ajout de termes d'erreurs permet de lever ce problème en introduisant de nouveaux aléas.

Par ailleurs, la présence des termes d'erreur amènera une discussion, lors de la présentation des expériences numériques, à propos des comportements de nos estimateurs sous différents régimes d'erreur ; nous pourrons ainsi expliquer, du moins en partie, les différences notables entre les valeurs prises par les estimateurs sur les données historiques à la fin de la partie précédente.

Notre objectif est d'estimer ϑ et les trajectoires de σ et $\overline{\sigma}$ grâce aux estimateurs de la partie précédente.

Cette démarche est à envisager dans le cadre plus large où, sur un intervalle de temps fixé, l'on n'observe pas directement un processus de diffusion X, mais $Y_t = X_t + \chi_t$, où χ_t est un bruit qui peut revêtir des formes très variées. La structure de notre problème est mathématiquement proche de celle de l'estimation en présence de *bruit de microstructure*. Cependant, nous soulignons le fait que notre démarche est complètement différente et que nos motivations ne sont pas celles des travaux sur la microstructure.

En 2001, Gloter et Jacod [36, 37] ont examiné un modèle simple où les observations sont de la forme $Y_{t_i} = X_{t_i} + \sqrt{\rho_n} U_i$, où les U_i sont des variables i.i.d. de loi $\mathcal{N}(0,1)$ indépendantes de la diffusion X. Ils ont établi une propriété de normalité asymptotique locale (LAN) pour l'estimation d'un paramètre dans la volatilité, en considérant différentes asymptotiques pour ρ_n. Des estimateurs sont également construits. Les vitesses d'estimation obtenues sont $\sqrt{n}$ si $n\rho_n \to u \in [0,\infty)$ et $(n/\rho_n)^{1/4}$ si $n\rho_n \to \infty$ avec $\sup_n \rho_n < \infty$. Notamment, dans le cas où ρ_n est constant, ce qui correspond au bruit de microstructure, la vitesse optimale est $n^{1/4}$. Bandi et Russell [14] ont examiné un modèle où les bruits étaient additifs et gaussiens. Zhang *et al.* [77] se sont intéressés au problème de l'estimation de la volatilité intégrée en présence d'un bruit additif de microstructure, et ont réussi à l'estimer à la vitesse $n^{1/6}$. Par la suite, Zhang [76] a réussi à estimer la volatilité intégrée à la vitesse $n^{1/4}$, optimale d'après les travaux de Gloter et Jacod. Ces deux travaux ont donné naissance à une famille d'estimateurs à plusieurs échelles. Une référence intéressante est le travail de Bibinger et Reiß [21] où des estimateurs sont obtenus à l'aide d'une équivalence asymptotique du modèle statistique sous-jacent avec un certain modèle de bruit blanc. Aït-Sahalia *et al.* [4] ont déterminé qu'il valait mieux utiliser toutes les données et bien modéliser le bruit plutôt que de choisir des pas d'échantillonnage plus larges et parfois arbitraires (voir la discussion à ce propos dans l'introduction de leur travail). D'autres formes d'erreur ont été considérées, comme des erreurs additives ayant une structure de dépendance dans Aït-Sahalia *et al.* [5] ou des erreurs d'arrondi pouvant être asymptotiquement petites, par Delattre et Jacod [27] ainsi que par Rosenbaum [69].

Les années suivantes ont vu le développement de la méthode dite de *pre-averaging* par Podolskij et Vetter [66] puis Jacod *et al.* [50], pour pouvoir estimer des fonctionnelles de la diffusion en présence d'un bruit pouvant appartenir à une classe très large de variables aléatoires, et exhibant potentiellement une dépendance en le processus X. Un grand nombre de résultats sont présentés dans le 16^e chapitre du livre de Jacod et Protter [52].

L'estimation non paramétrique des trajectoires du coefficient de volatilité à partir de données bruitées a également été l'objet de travaux ; Munk et Schmidt-Hieber [59] ont calculé des bornes inférieures pour les vitesses de convergence, en présence de bruit de microstructure et quand la volatilité est une fonction déterministe du temps. Munk et Schmidt-Hieber [60] ont ensuite donné des estimateurs non paramétriques dans ce même cadre. Reiß [67] a montré, toujours pour une fonction de volatilité déterministe, l'équivalence asymptotique avec une expérience de bruit blanc gaussien. Hoffmann *et al.* [46] ont étendu les possibilités d'estimation à une spécification stochastique de la volatilité, à l'aide de méthodes d'ondelettes.

La possibilité d'estimer efficacement la volatilité intégrée à partir de données bruitées a été étudiée également par Bibinger *et al.* [20], et plus récemment Jacod et Mykland [51] ainsi qu'Altmeyer et Bibinger [13].

Même si notre formalisme en est proche, nous ne serons pas dans un cadre de bruit de microstructure, car nous serons amenés à considérer des bruits qui ne sont pas asymptotiquement plus grands que le processus d'intérêt X. Cependant, nous traiterons de questions d'estimation de paramètres réels dans la volatilité, et d'estimation non paramétrique de réalisations de processus de volatilité, quand les données comportent un bruit. L'éventail de méthodes utilisées dans les références ci-dessus montre qu'utiliser des approximations de la volatilité intégrée, comme nous le faisons dans la partie précédente, n'est pas la solution permettant d'avoir les meilleures propriétés d'estimation quand les bruits sont importants.

Mais ici, nous sommes dans un cadre conceptuellement différent de celui des travaux sur la microstructure. Pour conserver les estimateurs développés précédemment, motivés par les interrogations laissées par les tests sur données réelles réalisés dans la première partie, nous nous plaçons dans une situation où les estimateurs restent convergents mais sont tout de même affectés par les erreurs.

Enfin, nous introduisons des erreurs en tant qu'*erreurs de modèle*. Il ne s'agit pas d'erreurs de mesure, mais bien d'un moyen de combler l'espace entre notre modèle mathématique et les données réelles, auxquelles le modèle n'est bien entendu pas parfaitement adapté.

Les questions qui nous préoccupent sont les suivantes :

Question 1 Les estimateurs introduits précédemment sont-ils robustes à l'ajout des erreurs de modèle décrites plus haut ?

Question 2 Quand la présence d'erreurs de modèle détériore les propriétés des estimateurs, notamment en termes de vitesse de convergence, existe-t-il des alternatives simples pour retrouver des estimateurs convergeant à la vitesse standard ?

Question 3 Dans notre modèle statistique ainsi enrichi, parvenons-nous à reproduire par la simulation les écarts constatés entre les valeurs des estimateurs dans les tests sur données réelles, et à cette fin, la spécification $\beta \geq 1/2$ n'est-elle pas trop restrictive ?

La question 3 est centrale, car sa réponse permettrait d'expliquer que des estimateurs différents puissent se concentrer autour de valeurs distinctes, ce que la considération du seul modèle de diffusion ne permet pas de faire.

Nous commençons par établir les propriétés des estimateurs $\hat{\vartheta}_{2,n}$ et $\hat{\vartheta}_{d,n}$ en présence d'erreurs de modèle : ce sont les mêmes estimateurs que dans la première partie, mais ils sont maintenant calculés à partir des observations discrètes du processus Y et non plus à partir de X. Nous gardons cependant la même notation. Les résultats de la partie précédente sont utilisés pour cela, et pour établir la loi limite des estimateurs nous cherchons le terme asymptotiquement prédominant, qui est dicté par le paramètre β ; pour des valeurs de β suffisamment élevées, nous retrouvons les lois limites de la première partie. Pour des valeurs de β faibles, les termes d'erreur dictent la loi limite à eux seuls. Il existe une valeur de β pour laquelle les termes d'erreur et ceux qui sont liés à la diffusion X sont asymptotiquement du même ordre, et des termes liés à chacun d'eux apparaissent alors dans les lois limites. On établit ce qui suit dans le théorème 3.1.

Résultat 5. *Si $\beta > 1/2$, nous avons trois régimes concernant l'estimateur $\hat{\vartheta}_{2,n}$:*

1. *si $\beta \in (1/2, 3/4)$, $\Delta_n^{1-2\beta}(\hat{\vartheta}_{2,n} - \vartheta) \to M_{\vartheta,\beta}$ en probabilité, où $M_{\vartheta,\beta}$ est une variable aléatoire où apparaissent ι_1 et ι_2, qui régissent le comportement asymptotique des erreurs de modèle ;*
2. *si $\beta > 3/4$, $\Delta_n^{-1/2}(\hat{\vartheta}_{2,n} - \vartheta) \to \mathcal{N}(0, V_\vartheta(\sigma, \overline{\sigma}))$ stablement en loi : nous retrouvons la même limite que dans la première partie ;*
3. *si $\beta = 3/4$, $\Delta_n^{-1/2}(\hat{\vartheta}_{2,n} - \vartheta) \to M_{\vartheta,\beta} + \mathcal{N}(0, V_\vartheta(\sigma, \overline{\sigma}))$ stablement en loi.*

Concernant les estimateurs $\hat{\vartheta}_{d,n}$, l'on observe un comportement similaire, décrit dans le théorème 3.2.

Résultat 6. *Si $\beta > 1/2$, nous avons les régimes suivants :*

1. *si $\beta \in (1/2, 1)$, $\Delta_n^{1-2\beta}(\hat{\vartheta}_{d,n} - \vartheta)$ converge en probabilité vers une variable aléatoire dépendant de $\iota_1, ..., \iota_d$;*
2. *si $\beta > 1$, $\Delta_n^{-1}(\hat{\vartheta}_{d,n} - \vartheta)$ converge en probabilité vers la même limite que dans la première partie.*

Dans chacun de ces résultats, une situation intéressante est celle où les termes liés à X et ceux qui correspondent aux erreurs de modèle ont asymptotiquement la même taille lorsque l'on cherche à établir la loi limite. Pour $d = 2$, la convergence stable en loi de $\Delta_n^{-1/2}(\hat{\vartheta}_{2,n} - \vartheta)$ de la partie précédente permet de conclure. Pour $d > 2$, ce n'est pas le cas et ainsi, dans le cas central $\beta = 1$ du résultat 6, nous ne sommes pas parvenus à conclure en utilisant un résultat comme le théorème IX.7.28 de Jacod et Shiryaev [54], qui contient le matériel nécessaire pour établir le résultat dans les autres cas. Nous énonçons que dans le cas où les processus σ et $\overline{\sigma}$ sont déterministes, $\Delta_n^{-1}(\hat{\vartheta}_{d,n} - \vartheta)$ converge en distribution vers une loi normale non centrée.

Nous introduisons ensuite un estimateur $\bar{\vartheta}_{3,n}$ qui, lorsque $d = 3$, converge vers ϑ lorsque $\beta \geq 1/2$ et nous établissons sa loi limite à l'aide du lemme 3.7 de Jacod [49], qui donne un théorème de convergence stable en loi pour des variables dépendantes. Nous avons le résultat suivant, énoncé dans le théorème 3.3.

Résultat 7. *En supposant que $\beta > 1/2$ et que*

$$\mathbb{E}\big[|\sigma_t^2 - \sigma_s^2|^2\big] + \mathbb{E}\big[|\overline{\sigma}_t^2 - \overline{\sigma}_s^2|^2\big] \leq c|t-s|^{2\alpha},$$

avec $c > 0$, $\alpha > 1/2$, et que les processus σ et $\overline{\sigma}$ sont presque-sûrement bornés inférieurement par une constante $\tilde{c} > 0$, nous avons la convergence

$$\Delta_n^{-1/2}(\bar{\vartheta}_{3,n} - \vartheta) \to \mathcal{N}(0, V_{\vartheta,3}(\sigma, \overline{\sigma}))$$

stablement en loi, où $\mathcal{N}(0, V_{\vartheta,3}(\sigma, \overline{\sigma}))$ est une variable aléatoire qui, conditionnellement à $\mathcal{F}$, est gaussienne et centrée. L'estimateur converge toutefois vers ϑ même si $\beta = 1/2$.

Ce résultat nous permet donc d'estimer ϑ même quand $\beta = 1/2$, et converge à la vitesse usuelle sans biais asymptotique. Nous pouvons ainsi répondre favorablement à la question 2 en ce qui concerne l'estimation de ϑ.

Pour ce qui est de l'estimation non paramétrique, nous commençons par établir dans le théorème 3.4 que lorsque $\beta > 1/2$, les estimateurs non paramétriques du modèle sans erreurs estiment σ_t^2 et $\overline{\sigma}_t^2$ à la vitesse $\Delta_n^{-(\frac{\alpha}{2\alpha+1} \wedge (2\beta-1))}$, qui est sous-optimale pour des valeurs de β trop petites.

Cela nous pousse à proposer de nouveaux estimateurs $\widetilde{\sigma}_{3,n,t}^2$ et $\widetilde{\overline{\sigma}}_{3,n,t}^2$, nécessitant l'utilisation préalable de $\bar{\vartheta}_{3,n}$. Le résultat relatif à ces estimateurs, prouvé dans le théorème 3.3 d'une manière standard, est le suivant.

Résultat 8. *Sous les mêmes conditions que celles du résultat 7, la suite*

$$\Big(\Delta_n^{-\alpha/(2\alpha+1)}\Big[\big|\widehat{\sigma}_{3,n,t}^2 - \sigma_t^2\big| + \big|\widehat{\overline{\sigma}}_{3,n,t}^2 - \overline{\sigma}_t^2\big|\Big]\Big)_{n\geq 1}$$

est tendue en probabilité, uniformément pour t dans un compact $\mathcal{D}$ inclus dans $(0, T]$.

Finalement, en reproduisant la preuve du théorème 2.4, nous montrons dans le théorème 3.6 qu'il est possible d'estimer ϑ à la vitesse $\Delta_n^{-1/2}$ et en atteignant la borne de Cramér-Rao (dans le sous-modèle avec des coefficients de volatilité déterministes), dès lors que $\alpha > 1/2$, $\beta > 3/4$ et $\sigma, \overline{\sigma}$ sont presque-sûrement bornés inférieurement.

Au vu de ces résultats, nous répondons aux questions 1 et 2 : les estimateurs de la première partie sont tous affectés par la présence d'erreurs, même dans le cas plutôt favorable $\beta \geq 1/2$. Nous sommes cependant parvenus à estimer ϑ et les processus de volatilité σ^2 et $\overline{\sigma}^2$ avec les vitesses usuelles en introduisant de nouveaux estimateurs, pourvu que l'on ait $d \geq 3$.

Dans les tests numériques, en appliquant nos estimateurs aux données historiques du marché à terme français d'électricité, nous obtenons des différences significatives entre les valeurs prises par $\hat{\vartheta}_{2,n}$ et $\bar{\vartheta}_{3,n}$ d'une part, et celles de $\hat{\vartheta}_{3,n}, ..., \hat{\vartheta}_{6,n}$ d'autre part ; ces dernières sont en moyenne six à dix fois plus faibles. Nous conduisons une expérience sur données simulées pour comprendre cela.

Cette expérience consiste à simuler des jeux de données pour différentes valeurs de ϑ et pour des termes d'erreur de modèle $\kappa_j^n \epsilon_i^j$ où les ϵ_i^j sont des variables i.i.d. de loi $\mathcal{N}(0,1)$ et κ_j^n est donné par $\frac{10^{-\chi}}{n^\beta}$, pour $\chi = 1, 2$ et pour β allant de 0,5 à 1,5. Dans chacune des configurations de simulation ainsi définies, nous réalisons 100 000 simulations et obtenons ainsi autant d'estimateurs de ϑ. À ϑ et χ fixés, nous regardons l'évolution de la moyenne et des quantiles des 100 000 estimateurs avec β. Nous obtenons par exemple, pour $\chi = 1$ et $\vartheta = 26,065$ a^{-1}, la figure 1.2.

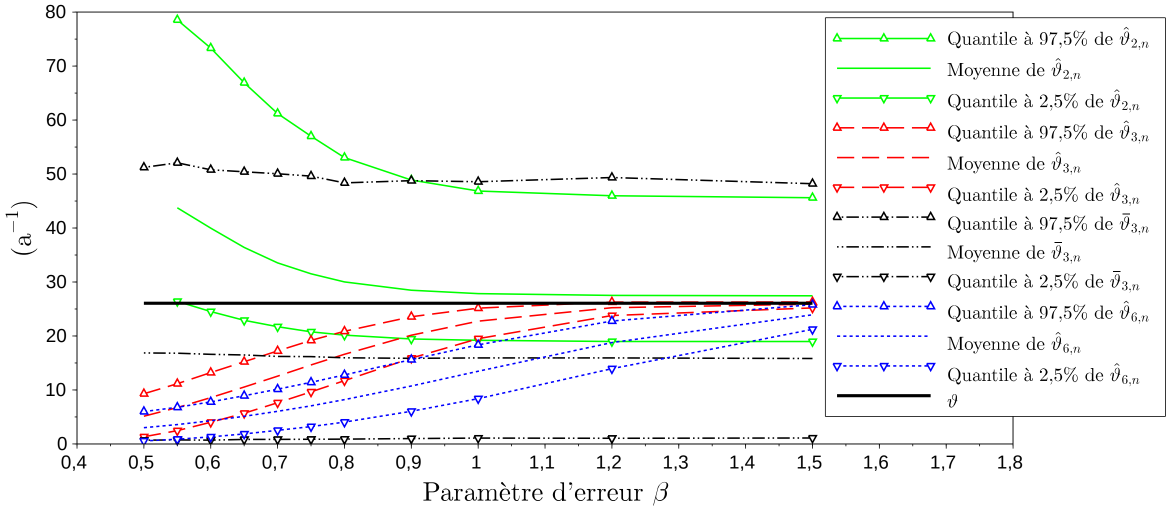

FIGURE 1.2 – Évolution de la moyenne et des quantiles des différents estimateurs avec β, pour $\chi = 1$ et $\vartheta = 26,065$ a^{-1}

La vraie valeur de ϑ est représentée en trait noir épais. Pour cette configuration, nous pouvons voir que les estimateurs $\bar{\vartheta}_{3,n}$ (en noir) sont relativement insensibles à la diminution de β, et que les estimateurs $\hat{\vartheta}_{2,n}$ (en vert) ont tendance à croître. Les estimateurs $\hat{\vartheta}_{3,n}$ (en rouge) et $\hat{\vartheta}_{6,n}$ (en bleu), aux performances correctes pour β grand, sont en revanche fortement biaisés pour β petit, sans que leur dispersion ne croisse beaucoup. Ceci est une configuration d'erreur de modèle dans laquelle un rapport cinq peut être observé entre deux estimateurs différents (par exemple, entre $\hat{\vartheta}_{2,n}$ et $\hat{\vartheta}_{6,n}$ pour $\beta = 0,7$). Un tel constat est un début d'explication aux valeurs très différentes prises par les estimateurs sur les données réelles.

Dans certaines configurations, nous arrivons à reproduire les écarts constatés entre les différents estimateurs sur les données réelles, ce qui permet de répondre favorablement à la question 3. Il n'est cependant pas évident, à partir de ces expériences, d'indiquer si la spécification $\beta < 1/2$ peut être exclue ou non pour le jeu de données réelles.

1.4 Troisième partie
Exécution optimale sur le marché de l'électricité

Ce travail a fait l'objet d'un article accepté pour publication dans *Mathematics and Financial Economics* [2].

L'on y considère un producteur tenu de mettre à la disposition d'un ensemble de consommateurs, à la date T, une certaine quantité d'électricité inconnue avant T. Sur le marché *day-ahead* que nous décrivons dans la section 1.5, il s'est déjà engagé à en livrer une partie à un certain coût. La différence entre la demande en T et cette quantité est appelée *prévision de la demande résiduelle en T, vue de $t = 0$*, notée D_0. Au fil de la période $[0, T]$, dont la durée peut aller de 8h45 à 31h45, le producteur va apprendre que la demande de ses consommateurs en T évolue d'une manière différente de celle qu'il avait prévue, ou que les aléas météorologiques vont changer la quantité d'énergie provenant des énergies renouvelables qu'il est capable de fournir, et sur laquelle il s'était engagé lors de l'enchère *day-ahead*. La prévision de la demande résiduelle en T vue de la date $t \in [0, T]$ est notée D_t, nous supposons qu'elle coïncide en $t = T$ avec la véritable demande résiduelle en T, et qu'elle suit la dynamique

$$dD_t = \mu dt + \sigma_d dB_t + \delta^+ dN_t^+ + \delta^- dN_t^-, \tag{1.6}$$

où $\mu \in \mathbb{R}$, $\sigma_d > 0$, $\delta^+ > 0$ et $\delta^- < 0$ sont des paramètres supposés connus, B est un mouvement brownien sur un certain espace filtré $(\Omega, \mathcal{F}, (\mathcal{F}_t)_t, \mathbb{P})$, et N^+, N^- sont deux processus de Poisson d'intensités λp^+ et λp^- , pour certains $\lambda > 0$, $p^+ \in [0, 1]$ et $p^- = 1 - p^+$. Tout se passe comme si des sauts avaient lieu à chaque fois qu'un processus de Poisson d'intensité λ sautait, et qu'ils étaient alors de hauteur π^+ avec la probabilité p^+ et de hauteur π^- avec la probabilité p^-. Une évolution déterministe de la demande résiduelle est donc anticipée, mais des modifications surviennent également *via* un mouvement brownien et des composantes de saut pur.

Entre les dates 0 et T, le producteur peut acheter ou vendre, en temps continu, des contrats assurant la livraison d'électricité à la date T pour l'aider à atteindre la demande résiduelle D_T. Sa vitesse d'achat/vente est donnée par $q = (q_t)_{t\in[0,T]}$, et son inventaire sur le marché est donné à la date t par

$$X_t = X_0 + \int_0^t q_s ds.$$

Par ailleurs, le prix auquel il peut effectuer une transaction q_t à la date t est donné par $P_t(q) = Y_t + \gamma q_t$, où $\gamma > 0$ est un facteur d'impact temporaire au sens d'Almgren et Chriss [12], et Y_t est le prix coté, supposé observable, sur le marché infrajournalier à l'instant t. Il a pour dynamique

$$dY_t = \nu q_t dt + \sigma_0 dW_t + \pi^+ dN_t^+ + \pi^- dN_t^-, \tag{1.7}$$

où $\nu > 0$ est un facteur d'impact permanent au sens d'Almgren et Chriss, $\sigma_0 > 0$, $\pi^+ > 0$ et $\pi^- < 0$. W est un mouvement brownien, corrélé avec B (qui conduit la demande résiduelle) à hauteur de $\rho \in (-1, 1)$. Les processus des sauts sont les mêmes que ceux de la demande, car on suppose que les informations conduisant à un saut de la demande résiduelle sont publiques et se répercutent donc sur le marché infrajournalier. Notons qu'en observant les données de marché, il existe des signes que certaines transactions exercent un impact parfois significatif. Par exemple, la figure 1.3 montre les transactions ayant eu lieu sur le marché allemand (qui fonctionne exactement comme le marché français) pour des contrats assurant une livraison le 16 décembre 2010 à 7h. Chaque point bleu correspond à une transaction. Le carré vert représente le prix issu de l'enchère *day-ahead*. L'on peut constater que les premières transactions ont lieu à des prix proches de ce prix *spot*, puis qu'aux alentours de 4h30, plusieurs transactions interviennent de manière rapprochée, ce qui a pour conséquence de faire monter le prix de manière considérable.

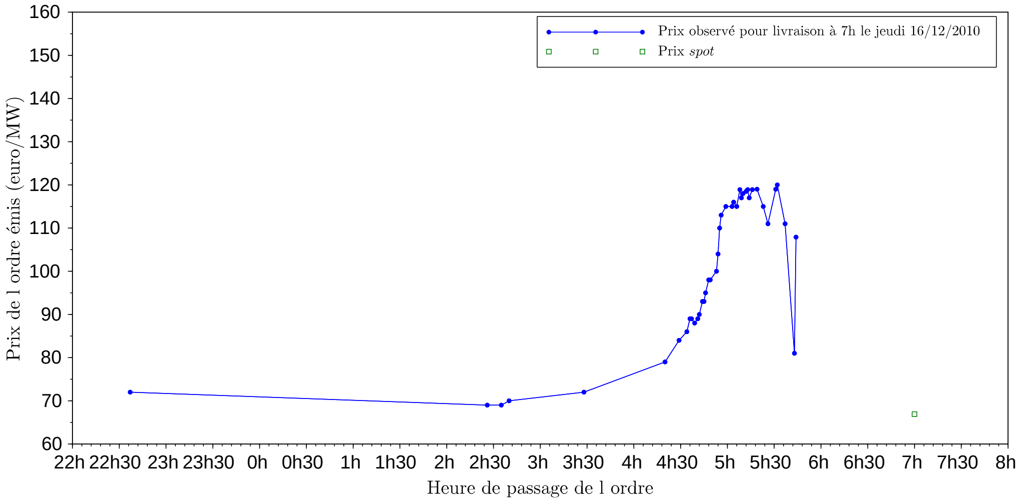

FIGURE 1.3 – Prix des transactions sur le marché infrajournalier allemand pour livraison le 16 décembre 2010 à 7h

À la date $T - h$, où $h \geq 0$ est un certain délai, le producteur peut décider de mobiliser des unités de production non prévues lors de l'enchère *day-ahead*, qui pourront lui fournir en T une quantité non aléatoire $\xi \geq 0$ au coût de production $c(\xi)$.

Enfin, en T, le régulateur compare la quantité mise à disposition $X_T + \xi$ à la demande D_T, et pénalise le producteur à hauteur de $\frac{\eta}{2}(D_T - X_T - \xi)^2$, où $\eta > 0$, car il déséquilibre le réseau dans sa globalité s'il n'est pas en mesure de réaliser l'équilibre production/consommation sur son périmètre.

Le problème que résout le producteur est donc

$$\min_{q \in \mathcal{A},\, \xi \in L^0_+(\mathcal{F}_{T-h})} \mathbb{E}\Big[\int_0^T q_t P_t(q) dt + c(\xi) + \frac{\eta}{2}(X_T + \xi - D_T)^2 \Big],$$

où $L^0_+(\mathcal{F}_t)$ est l'ensemble des variables aléatoires $\mathcal{F}_t$-mesurables et positives, et $\mathcal{A}$ est un ensemble de contrôles à préciser.

Ce problème peut être étudié à l'aide des nombreux travaux traitant d'exécution optimale en présence d'impact de marché. En 1998, Bertsimas et Lo [17] ont utilisé, pour la première fois et dans un modèle en temps discret, la programmation dynamique pour traiter un problème d'exécution optimale. Leur objectif était de minimiser l'espérance du coût des T transactions à effectuer avant la date finale afin d'acquérir une quantité d'actifs fixée. En 2000, Almgren et Chriss [12], dans un article fondateur, ont introduit un modèle d'impact de marché dans lequel chaque transaction donne lieu à une modification de prix temporaire et à une autre qui est permanente. Leur objectif était de définir une stratégie de liquidation en temps discret tout en limitant l'espérance des pertes liées aux coûts de transaction et au risque de volatilité. Le succès de ce modèle réside autant dans sa simplicité, *via* la considération de fonctions d'impact linéaires permettant souvent la détermination de stratégies explicites, que dans la richesse de la modélisation qu'il permet d'effectuer.

Par la suite, de nombreuses discussions autour de ce modèle d'impact ont eu lieu. Almgren [10] argumente en faveur d'une fonction d'impact temporaire non linéaire, et Almgren [11] présente un modèle où la liquidité et la volatilité sont rendues stochastiques mais sans impact permanent. Forsyth *et al.* [29] se fixent un critère de type moyenne–variance, et retrouvent les résultats d'Almgren et Chriss si le prix non affecté est un mouvement brownien arithmétique.

Assorties de considérations empiriques, des argumentations favorables à la modélisation d'un impact *transient*, c'est-à-dire avec un effet s'estompant avec le temps, se sont développées. Obizhaeva et Wang [63] ont motivé ce choix, et ont par ailleurs modélisé la dynamique des carnets d'ordres limites, sans se limiter à leurs propriétés statiques. Les auteurs suggèrent d'utiliser des ordres continus et discrets. Alfonsi *et al.* [6] ont étendu ce modèle, en temps discret, à des temps d'action libres et en considérant que la vitesse de résilience du carnet d'ordres peut varier au cours du temps. Les mêmes auteurs ont également repris le modèle d'Obizhaeva et Wang dans Alfonsi *et al.* [7], en supposant que le carnet d'ordres peut avoir une forme quelconque. Weiss [74] a étendu ce modèle en faisant dépendre la vitesse de résilience de la taille de l'ordre passé. Cont *et al.* [25] ont modélisé le carnet d'ordres et la manière dont les événements qui y surviennent impactent le prix. Schied et Schöneborn [70] n'ont pas utilisé le paradigme de la maximisation de l'espérance des coûts, mais ont opté pour la minimisation d'une fonction d'utilité. De plus, leur problème est différent de la liquidation optimale en temps fini, car ils se placent dans un horizon temporel infini.

Les modèles d'impact ont également été examinés du point de vue des manipulations de marché qu'ils rendent possibles ; par exemple, il n'est pas souhaitable qu'un acteur du marché puisse, à partir d'une richesse initiale, passer des ordres d'achat et de vente puis revenir à sa richesse de départ à l'aide d'une stratégie dont l'espérance du coût est négative. Ces irrégularités de marché sont explicitées et testées dans, notamment, les travaux de Huberman et Stanzl [47], Alfonsi *et al.* [8], Gatheral [31], Gatheral *et al.* [32] et Alfonsi *et al.* [9].

D'un point de vue méthodologique, nous mettons en avant le travail récent de Chen *et al.* [23], où les auteurs recherchent la stratégie optimale d'achat dans un modèle où la profondeur de marché est stochastique et le carnet d'ordres est résilient. Leur approche est fondée sur la programmation dynamique. Alors que le modèle ne permettait pas d'obtenir une solution analytique, les auteurs ont pu traiter deux problèmes différents, mais proches, en considérant un espace de contrôles plus restreint et un autre plus large. Ils obtiennent ainsi deux bornes pour la fonction valeur du problème initial, et montrent qu'elles sont peu éloignées l'une de l'autre.

À propos de problèmes plus proches de celui qui va nous préoccuper, Henriot [41] a étudié la manière dont le marché infrajournalier peut aider un producteur à compenser les aléas qu'il encourt du fait de la difficulté de prévoir sa production éolienne future, même à court terme. Garnier et Madlener [30] ont discuté de l'opportunité, pour un producteur dont le rendement des installations est aléatoire, d'entrer sur le marché infrajournalier à un instant donné ou de reporter cette entrée à plus tard. Pour des références quant aux problèmes posés par les difficultés de prévision de la production éolienne, l'on peut consulter les travaux de Bensoussan *et al.* [15] et Giebel *et al.* [35].

En plus d'être fondé sur le modèle d'Almgren et Chriss, dont nous avons cité plusieurs études et extensions précédemment, notre problème partage des similarités avec les questions d'exécution optimale. La différence principale réside dans le fait que la cible est ici aléatoire ; il ne s'agit pas d'acheter ou de vendre une quantité d'actifs fixée à l'avance, mais de poursuivre la cible D_t en horizon de temps fini.

Nous cherchons ici à répondre aux questions suivantes :

Question 1 Peut-on décrire, de façon explicite si possible, la stratégie que doit suivre le producteur sur le marché infrajournalier et son choix de production ξ pour minimiser ses coûts moyens ?

Question 2 Dans quelle mesure ce modèle nous permet-il d'étendre les propriétés de celui d'Almgren et Chriss, jusqu'ici ordinairement utilisé pour des problèmes d'exécution optimale en présence d'impact de marché sur un marché financier classique ?

Dans un premier temps, nous traitons le problème simplifié où $h = 0$ et où il n'y a pas de sauts dans les dynamiques de la demande résiduelle et du prix. Le problème admet une écriture simple permettant de le traiter dans un cadre de programmation dynamique. Soit

$$v(t, x, y, d) := \inf_{q \in \mathcal{A}_t, \xi \in L^0_+(\mathcal{F}_T)} J(t, x, y, d; q, \xi), \tag{1.8}$$

avec

$$J(t, x, y, d; q, \xi) := \mathbb{E}\Big[\int_t^T q_s(Y_s^{t,y} + \gamma q_s) ds + c(\xi) + \frac{\eta}{2}(X_T^{t,x} + \xi - D_T^{t,d})^2 \Big],$$

où les notations $X_s^{t,x}$, $Y_s^{t,y}$ et $D_s^{t,d}$ représentent les valeurs à la date $s > t$ des processus X, Y et D respectivement partis des valeurs x, y et d à l'instant t. De plus, $\mathcal{A}_t$ est l'ensemble des processus réels $q = (q_s)_{t \leq s \leq T}$ tels que q_s est $\mathcal{F}_s$-adapté et $\mathbb{E}\big[\int_t^T q_s^2 ds \big] < \infty$.

Ce problème a vocation à être traité de manière rétrograde. Notamment, l'on commence par déterminer que le choix optimal de ξ, noté $\xi_T^*(X_T, D_T)$, est obtenu à l'instant terminal

T en résolvant un problème d'optimisation :

$$\xi_T^*(X_T, D_T) = \arg\min_{\xi \geq 0} \left[c(\xi) + \frac{\eta}{2}(X_T + \xi - D_T)^2 \right] = \frac{\eta}{\eta + \beta}(D_T - X_T)\mathbf{1}_{D_T - X_T \geq 0},$$

pour la fonction de coût $c(\xi) = \frac{\beta}{2}\xi^2$. Ainsi, la condition terminale de notre problème est

$$\begin{aligned} v(T, x, y, d) =& c(\xi_T^*(x, d)) + \frac{\eta}{2}(x + \xi_T^*(x, d) - d)^2 \\ =& \frac{1}{2}\frac{\eta\beta}{\eta + \beta}(d - x)^2 \mathbf{1}_{d - x \geq 0} + \frac{\eta}{2}(d - x)^2 \mathbf{1}_{d - x < 0}. \end{aligned}$$

À cause des indicatrices, nous n'avons pas l'espoir d'obtenir une expression analytique pour la fonction valeur et la stratégie optimale. Dans l'esprit du travail de Chen *et al.*, nous élargissons l'espace des contrôles en autorisant ξ à prendre des valeurs négatives, même si cela n'a pas de sens pour une quantité d'électricité produite. Avec cet élargissement, nous obtenons une valeur terminale régulière, puis nous calculons la fonction valeur et la stratégie optimale *via* l'approche par équations aux dérivées partielles (EDP), en résolvant l'EDP de Hamilton-Jacobi-Bellman (HJB) de manière explicite. Nous obtenons notamment la fonction valeur et la stratégie optimale dans le problème approché dans le théorème 4.1, et une propriété de martingale dans la proposition 4.1.

Résultat 9. *Dans le problème approché sans sauts et sans délai, la stratégie optimale est donnée par*

$$\hat{q}_s = \frac{r(\eta, \beta)(\mu(T - s) + (D_s^{t,d} - \hat{X}_s^{t,x,y,d})) - \hat{Y}_s^{t,x,y,d}}{(r(\eta, \beta) + \nu)(T - s) + 2\gamma},$$

où $r(\eta, \beta) = \frac{\eta\beta}{\eta+\beta}$ *et* $\hat{X}_s^{t,x,y,d}$, $\hat{Y}_s^{t,x,y,d}$ *sont les valeurs en* $s \geq t$ *des processus* X *et* Y *contrôlés optimalement depuis l'état* (t, x, y, d)*. De plus,*

le processus $(\hat{q}_s)_{t \leq s \leq T}$ *est une martingale.*

Cette propriété de martingale est intéressante et doit être mise en regard des résultats obtenus par Almgren et Chriss [12] ; dans leur modèle où la cible est constante, la vitesse d'achat/vente est constante. Ici, la propriété de martingale de la cible est transférée à la trajectoire optimale d'achat/vente. Il s'agit d'un élément de réponse à la question 2.

Comme nous avons dû faire l'hypothèse que ξ pouvait être négatif pour établir ce résultat, nous avons besoin de bornes pour la différence entre la fonction valeur trouvée dans le théorème 4.1 et la vraie fonction valeur du problème (1.8), qui est inconnue. Nous parvenons à obtenir de telles bornes, exprimées en fonction des paramètres x, y, d à l'instant initial, et à donner des conditions sous lesquelles elles sont petites.

Nous incorporons les sauts dans les dynamiques de la demande résiduelle et du prix coté sur le marché infrajournalier, qui sont donc données par (1.6) et (1.7). Nous suivons la même démarche que précédemment, en relâchant la contrainte de positivité de ξ, ce qui permet d'obtenir la fonction valeur et la stratégie optimale dans le problème approché, *via* la résolution de l'équation de HJB. Ces résultats sont donnés dans le théorème 4.2, et font apparaître que le producteur cherche à tirer profit des éventuels sauts pouvant se produire en achetant puis en revendant (dans le cas où des sauts positifs de prix sont les plus probables) des quantités de contrats potentiellement importantes. Puis nous donnons des bornes sur la différence entre la fonction valeur approchée et la vraie fonction valeur. Cette fois-ci, les résultats relatifs à la stratégie optimale $(\hat{q}_s^{(\lambda)})_s$ du problème approché sont les suivants.

Résultat 10. *Dans le problème approché sans délai, la stratégie optimale est donnée par*

$$\hat{q}_s^{(\lambda)} = \hat{q}_s^{(0)} + \lambda \frac{r(\eta,\beta)\delta(T-s) + \frac{\pi}{4\gamma}(r(\eta,\beta)+\nu)(T-s)^2}{(r(\eta,\beta)+\nu)(T-s)+2\gamma}.$$

où $\pi = p^+\pi^+ + p^-\pi^-$ *et* $\delta = p^+\delta^+ + p^-\delta^-$ *sont respectivement les hauteurs moyennes des sauts de prix et de demande. De plus,*

$$\textit{le processus } \left(\hat{q}_s^{(\lambda)} + \frac{\lambda\pi}{2\gamma}(s-t)\right)_{t\le s\le T} \textit{ est une martingale.}$$

Il est intéressant de noter que la présence de sauts sur le prix induit que la stratégie optimale n'est plus martingale (*via* π), alors que le fait qu'ils soient présents dans la dynamique de la demande résiduelle n'implique pas la perte de cette propriété.

Nous traitons ensuite le problème avec délai de production. Il apparaît que le problème approché, où ξ n'est pas nécessairement positif, peut être découpé en deux sous-problèmes : après $T-h$, on doit résoudre le problème sans délai et sans capacité de production (il suffit de prendre $\beta \to \infty$ dans le problème précédent), et avant $T-h$, on résout le problème sans délai et avec capacité de production. Tous nos traitements du cas avec délai se limitent au cas sans sauts ; au prix d'une complexification notable des calculs, des résultats analogues peuvent cependant être établis en la présence de sauts.

Nous obtenons le résultat suivant dans le problème approché avec délai.

Résultat 11. *Vu de l'instant initial* 0*, la fonction valeur du problème approché et avec délai* h *est égale à la somme de la fonction valeur du problème approché sans délai de production, donnée dans le théorème 4.1, et d'un terme* K_h *dépendant de* h *et des paramètres du problème, mais pas de* T*,* x*,* y *ou* d*.*
De plus, la stratégie optimale $(\hat{q}_s^h)_s$ *du problème approché est une martingale pour* $s \in [T-h,T]$*, pour* $s \in [0,T-h]$*, mais aussi pour* $s \in [0,T]$*.*

Nous sommes ensuite en mesure de donner des bornes pour la différence entre la fonction valeur et la fonction valeur approchée, dans le cas avec délai.

Nous effectuons, dans chacun des trois problèmes considérés, des simulations de la dynamique des processus en jeu quand nous suivons la trajectoire optimale dans les modèles approchés. Par exemple, la figure 1.4 présente, pour un certain choix de paramètres avec notamment $T = 24$ h et $h = 4$ h, une trajectoire simulée du processus X et de la prévision de la demande résiduelle D au cours du temps, en l'absence de sauts. L'apparence presque constante de la pente de X est à rapprocher du fait que la vitesse d'achat/vente est une martingale. Observons le saut dans l'inventaire à la date $T-h$, correspondant au choix de production effectué, et qui nous engage pour la date T.

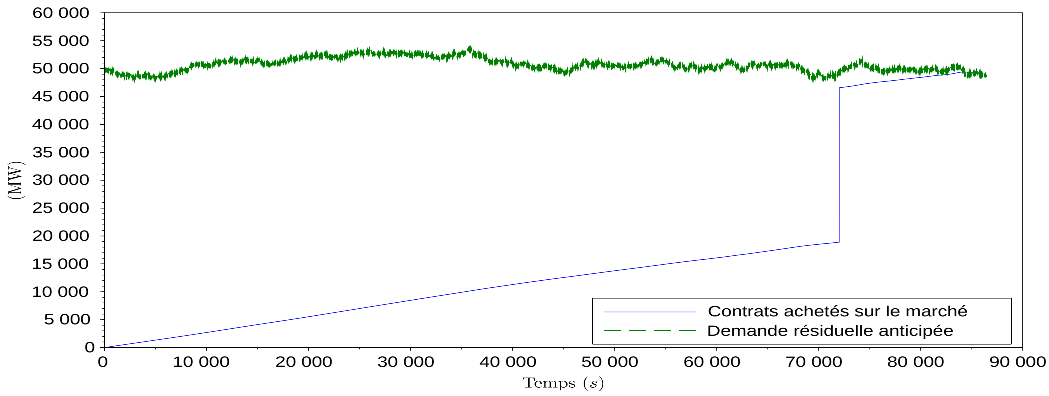

FIGURE 1.4 – Évolution de l'inventaire $\hat{X}_t + \xi \mathbf{1}_{t \geq T-h}$ (avec choix de production à l'instant $T-h$) et de la prévision de la demande résiduelle D_t

Finalement, pour répondre à la question 1, observons qu'il n'est pas possible de caractériser explicitement la fonction valeur et la stratégie optimale dans le problème général. Mais en considérant un problème approché, des formules analytiques sont obtenues et il est possible d'indiquer si la fonction valeur approchée est proche, ou non, de la fonction valeur du problème de départ.

Concernant la question 2, la principale observation est celle que nous avons formulée sur le caractère martingale de la stratégie optimale (corrigée d'un terme linéaire en présence de sauts sur les prix) dans les problèmes approchés. Il s'agit d'un pendant intéressant au caractère constant de la stratégie dans le cadre du problème d'exécution optimale classique, et la question de son extension à des problèmes plus généraux que le nôtre se pose.

Un autre problème se posant est celui de la pertinence de la stratégie issue du théorème 4.2 : comme nous l'avons relevé plus haut, celle-ci est une combinaison de la stratégie du cas sans sauts et d'une tentative de profiter des sauts pouvant se produire, en suivant une stratégie d'achat/vente dont l'ampleur n'est limitée que grâce à l'impact temporaire. La formulation du problème du producteur à l'aide d'un autre critère, afin de ne pas permettre de telles stratégies, serait une extension intéressante et immédiate de ce travail.

1.5 Description des marchés de gros de l'électricité

Nous présentons ici les marchés de gros de l'électricité en France ; en Europe, d'autres marchés semblables existent, et leur fonctionnement ne diffère que très peu.

Marché de contrats à terme

Fonctionnement du marché

Nous décrivons dans un premier temps le marché de contrats financiers sur l'électricité, géré en France par l'*European Energy Exchange (EEX)*. C'est un marché financier classique, de gré à gré, où l'on trouve des contrats « financiers », c'est-à-dire qu'ils ne donnent pas lieu à la livraison d'électricité (contrairement aux contrats « physiques » des marchés court-terme) même s'ils sont caractérisés par une période de livraison future. Ils peuvent être vus comme des produits dérivés, fondés sur le prix de la livraison d'un mégawatt-heure répartie sur cette période, juste avant le début de celle-ci. Étant donné que les contrats ne donnent pas lieu à une livraison, le marché n'a pas besoin d'être restreint aux acteurs ayant des capacités de production ou de réception d'électricité ; il est ouvert aux spéculateurs.

Chaque transaction dans un de ces contrats est caractérisée par une période de livraison (fictive, puisque celle-ci n'a pas lieu), un volume (nombre de contrats) et un prix. Ce sont des produits qui peuvent être échangés durant une période dite de cotation. *Via* des appels de marge quotidiens, un contrat donne finalement lieu à l'échange, entre son acheteur et son vendeur, de son prix d'achat contre le dernier prix coté lors de la période de cotation. Pour une description du marché à terme norvégien, dont le fonctionnement est assez voisin du marché français, on se référera à Bjerksund *et al.* [22, 58] et au travail de modélisation approfondi de Benth et Koekebakker [16].

Données disponibles relatives à ce marché

Dans le cas du marché français, les durées de livraison disponibles sont une journée, une semaine, un mois, un trimestre et une année. Nous donnons les produits existants dans les tableaux 1.1 et 1.2, en indiquant, à des fins d'illustrations, à quelles dates de livraison ils correspondaient exactement le 30 juin 2008 et le 1er juillet 2008. Les produits intitulés *j Month-ahead* sont relatifs à une livraison sur une période d'un mois, commençant le premier jour du j^{e} mois suivant le mois de cotation. De la même façon, les produits intitulés *j Quarter-ahead* sont relatifs à la livraison sur une période de trois mois, commençant le 1er janvier, le 1er avril, le 1er juillet ou le 1er octobre, suivant la valeur de j et la date de cotation. Le produit *1 Year-ahead* correspond à une livraison sur l'année civile suivant celle dans laquelle l'on se trouve au moment de la cotation. Alors qu'un *produit* est caractérisé par un temps restant avant livraison (*via* le chiffre au début de son nom) et désigne donc une période différente lorsque l'on passe d'un mois, d'un trimestre ou d'une année à l'autre, un *contrat* est associé à une période bien définie, et n'est coté que pendant un certain nombre de mois, pendant lesquels il coïncide avec l'un des *produits*.

Pour notre étude, nous avons disposé des données des produits *1–6 Month Ahead* sur le marché français, du 6 décembre 2001 au 30 décembre 2013. Chaque jour ouvré, et pour chacun des six produits, nous avons accès au dernier prix auquel une transaction a eu lieu. Ainsi, pour le contrat rattaché à la livraison (fictive) d'un mégawatt-heure au cours d'un mois donné, nous avons un prix pour chaque jour ouvré des six mois précédant la période de livraison, soit de l'ordre de 120 prix (6 mois comportant, chacun, environ 20 jours ouvrés).

Produit	Date de cotation : 30 juin 2008		
	Nom du contrat	Début de livraison	Fin de livraison
1 Month Ahead	Juillet 2008	01/07/2008	31/07/2008
2 Month Ahead	Août 2008	01/08/2008	31/08/2008
3 Month Ahead	Septembre 2008	01/09/2008	30/09/2008
4 Month Ahead	Octobre 2008	01/10/2008	31/10/2008
5 Month Ahead	Novembre 2008	01/11/2008	30/11/2008
6 Month Ahead	Décembre 2008	01/12/2008	31/12/2008
1 Quarter Ahead	3e trimestre 2008	01/07/2008	30/09/2008
2 Quarter Ahead	4e trimestre 2008	01/10/2008	31/12/2008
3 Quarter Ahead	1er trimestre 2009	01/01/2009	31/03/2009
1 Year Ahead	Année 2009	01/01/2009	31/12/2009

Table 1.1 – Contrats cotés le 30 juin 2008

Produit	Date de cotation : 1er juillet 2008		
	Nom du contrat	Début de livraison	Fin de livraison
1 Month Ahead	Août 2008	01/08/2008	31/08/2008
2 Month Ahead	Septembre 2008	01/09/2008	30/09/2008
3 Month Ahead	Octobre 2008	01/10/2008	31/10/2008
4 Month Ahead	Novembre 2008	01/11/2008	30/11/2008
5 Month Ahead	Décembre 2008	01/12/2008	31/12/2008
6 Month Ahead	Janvier 2009	01/01/2009	31/01/2009
1 Quarter Ahead	4e trimestre 2008	01/10/2008	31/12/2008
2 Quarter Ahead	1er trimestre 2009	01/01/2009	31/03/2009
3 Quarter Ahead	2e trimestre 2009	01/04/2009	30/06/2009
1 Year Ahead	Année 2009	01/01/2009	31/12/2009

Table 1.2 – Contrats cotés le 1er juillet 2008

Marchés court-terme

En France, deux marchés, gérés par *EPEX SPOT SE*, permettent d'échanger des contrats pour une livraison ayant lieu moins de trente-six heures dans le futur. Précisément, on découpe un jour j fixé en vingt-quatre périodes d'une heure, pour lesquelles il va être possible de s'engager à acheter ou à vendre de l'électricité. Le premier marché est appelé *day-ahead*, c'est une enchère à l'aveugle qui a lieu à midi le jour $j-1$, et au cours de laquelle un prix d'équilibre est fixé pour chaque heure du jour j, suivant les ordres d'achat et de vente soumis par les acteurs. Le prix d'équilibre est appelé *prix spot*, même s'il ne s'agit pas d'un prix instantané au sens des marchés financiers usuels. Il doit plutôt être vu comme un témoin de l'équilibre pour le jour j prévalant à $j-1$.

Le deuxième marché, dit *infrajournalier*, ouvre à quinze heures en $j-1$, après la publication des résultats de l'enchère *day-ahead* ; pour chacune des vingt-quatre périodes du jour

j, il est possible d'acheter ou de vendre des contrats donnant lieu à la livraison d'électricité pendant la période. Ces contrats sont caractérisés par un prix et un volume, exprimé en mégawatts. Chacun des contrats peut être échangé jusqu'à 45 minutes avant le début de la livraison. Pour chaque période, les ordres passés sont enregistrés dans un carnet d'ordre, et dès qu'un ordre d'achat et un ordre de vente se correspondent, ils sont exécutés.

La chronologie du fonctionnement du marché infrajournalier est représentée sur la figure 1.5. Elle illustre le chevauchement des périodes de cotation, et la disparition progressive des contrats jusqu'à 23h15 le jour j.

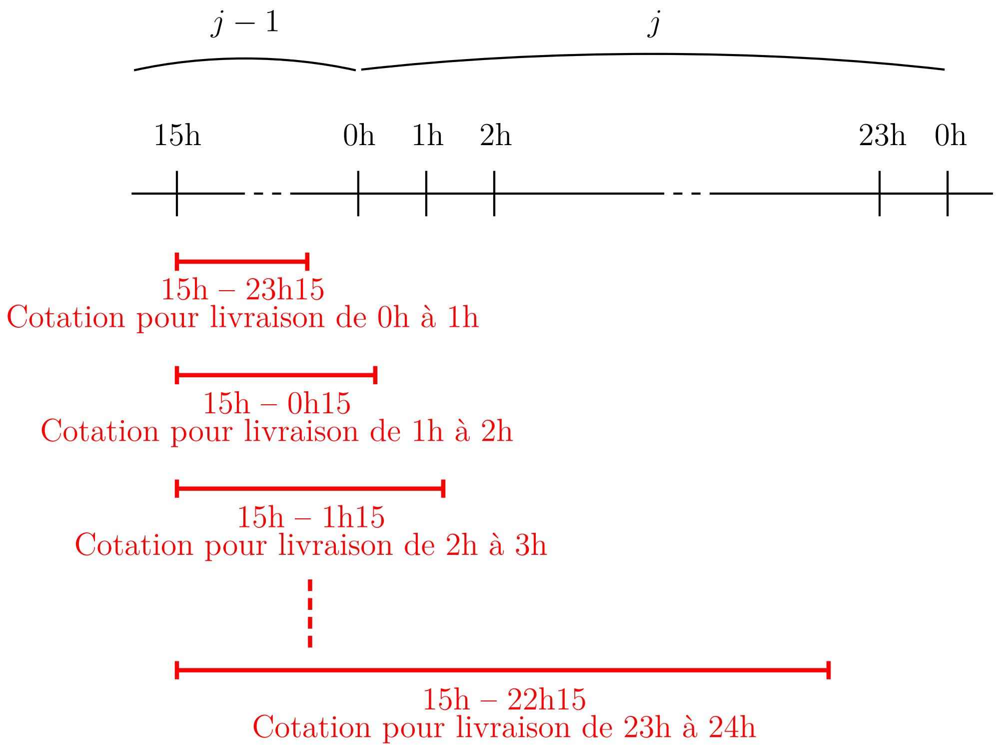

FIGURE 1.5 – Fonctionnement du marché infrajournalier

1.6 Perspectives

Nous réalisons ici une synthèse des travaux et donnons quelques perspectives de prolongement émergeant au vu de nos résultats.

Dans le problème d'estimation efficace dans la volatilité d'un processus de diffusion conduit par deux mouvements browniens, dont l'étude est conduite dans les chapitres 2 et 3, nous sommes parvenus à estimer les composantes paramétrique et non paramétrique du coefficient de volatilité, chacune avec la vitesse optimale de leur paradigme. Pour la composante paramétrique, l'estimation est réalisée efficacement, dans le sens où la borne de

Cramér-Rao est atteinte dans le sous-modèle où la spécification de la volatilité est entièrement déterministe. Nous avons mis en exergue une dégénérescence du modèle, typique des modèles de Heath-Jarrow-Morton, à laquelle nous avons paré grâce à l'ajout d'erreurs de modèle. Nous avons ensuite observé comment cet ajout d'erreurs modifiait le comportement des estimateurs, et constaté qu'il était possible de proposer des estimateurs alternatifs qui atteignent la vitesse optimale dans ce cadre. Nous sommes en outre parvenus à reproduire, par la simulation, les écarts importants constatés entre les valeurs fournies par nos différents estimateurs sur les données réelles ; ceci est un plaidoyer pour la prise en compte des termes d'erreur dans la modélisation.

Nous avons souligné que notre cadre de travail comportait des similitudes avec celui du traitement du bruit de microstructure, mais que l'esprit des travaux en est fondamentalement différent. Cela a justifié que nous continuions à utiliser des estimateurs fondés sur la variation quadratique réalisée, ce qui est à éviter quand les données comportent des bruits plus importants que dans notre modèle.

Une perspective est d'effectuer le rapprochement avec le bruit de microstructure pour avoir une spécification plus riche des termes d'erreur et utiliser les techniques les plus récentes de ce domaine de recherche, comme les méthodes de *pre-averaging* de Jacod *et al.* [50], pour effectuer des tests numériques plus poussés et examiner plus en profondeur la question de l'impact des erreurs sur l'estimation d'un paramètre fini-dimensionnel, dans notre modèle à deux facteurs.

Nous nous sommes également intéressés, dans le chapitre 4, à la détermination de la stratégie optimale d'un producteur, désireux d'acheter ou de vendre des contrats sur le marché infrajournalier de l'électricité pour réduire son exposition aux aléas de ses unités de production. Nous avons supposé que, par ses actions, le producteur exerçait un impact sur les prix du marché. Ayant le souci d'obtenir des formules analytiques à des fins d'interprétation économique, nous avons dû modifier légèrement le problème posé en élargissant l'ensemble des possibilités d'action du producteur. Nous avons obtenu, grâce à la résolution de l'équation aux dérivées partielles de Hamilton-Jacobi-Bellman, l'espérance des coûts du producteur et une caractérisation explicite de sa stratégie dans le problème approché. Nous avons pu établir que cette stratégie était une martingale (ou une sur/sous-martingale en présence de sauts sur les prix), ce qui étend de manière intéressante les résultats d'Almgren et Chriss [12] à la poursuite d'une cible aléatoire en présence d'impact de marché. Enfin, nous avons pu donner des conditions pour que la valeur et la stratégie du problème approché soient proches de celles du problème de départ, et nous avons quantifié cette proximité.

L'examen du problème approché montre que si le producteur anticipe que les prix cotés vont subir des sauts positifs (resp. négatifs), il adopte une stratégie visant à acheter puis à revendre (resp., à vendre puis à racheter) une grande quantité de contrats dans l'espoir que des sauts seront survenus et qu'il en tirera ainsi profit. Cette stratégie est difficilement compatible avec une optique industrielle. Une extension intéressante est donc la formulation d'un problème voisin, assimilable à de la *gestion de risques*, où il s'agirait de couvrir le coût de la satisfaction des consommateurs par les seuls moyens de production, à l'aide du marché infrajournalier. Ce problème pourrait alors être plongé dans le cadre plus général de la couverture en présence d'impact de marché.

Enfin, le modèle des chapitres 2 et 3 est adapté aux contrats à terme, et l'on en trouve également sur le marché infrajournalier, même si la distance à maturité y est bien plus faible. Une démarche intéressante, visant à unifier les deux domaines de recherche des présents travaux, serait d'enrichir le modèle brownien arithmétique considéré pour le prix dans le chapitre 4 en utilisant la diffusion à deux facteurs des chapitres précédents. L'objectif serait alors de réaliser l'estimation des composantes de la volatilité sur le marché infrajournalier, et d'observer comment sa qualité influe sur la pertinence de la stratégie d'achat/vente du producteur ayant connaissance des valeurs estimées.

Chapitre 2

Efficient estimation in a two-factor model from historical data

2.1 Introduction

2.1.1 Motivation

This chapter deals with estimation procedures for multidimensional diffusion processes, with a volatility structure including both parametric and nonparametric components. We care for efficient estimation of a scalar parameter in the volatility, in presence of nonparametric nuisance, while providing point estimates of nonparametric components too. The processes of interest follow the multiple Brownian factor representation, as in the Heath-Jarrow-Morton (HJM) framework for forward rates, for instance in Heath *et al.* [40], or for electricity forward contracts in Benth and Koekebakker [16].
In the context of interest rate models, some studies have focused on estimation in a 1-dimensional fully parametric diffusion context, see Aït-Sahalia [3] for an overview. Other assume a fully nonparametric setting, like in Aït-Sahalia [3] where observation times tend to infinity, the drift is supposed to be a linear function of the level of the process, and the volatility is estimated thanks to the Kolmogorov forward equations. The setting of Stanton [72] is similar to the one of Aït-Sahalia [3], but the hypothesis of linearity of the drift is relaxed. The authors of Jeffrey *et al.* [55] assume that the volatility is an unspecified function of the level of the process and of the distance to maturity. Bhar *et al.* [19] focus on maximum likelihood estimation in a HJM model, and Bhar and Chiarella [18] use nonlinear filtering to estimate parameters in a one-factor HJM model.
More generally, estimation of the volatility function of a (multidimensional) process observed over some period $[0, T]$ has been the subject of various works in the asymptotics where observation times tend to recover the whole period of observation. See for instance Genon-Catalot and Jacod [33] in a parametric and semiparametric setting while knowing the form of the volatility function, and Jacod [49] and the references therein in a nonparametric framework.

Our setting will be motivated by the context of prices of specific forward contracts, which are available on the electricity market. Interest rate models have been applied to the pricing of such contracts: see for instance Hinz *et al.* [42], in which an analogy between interest rate models and forward contracts prices models is performed, the maturity in the former framework being a date of delivery in the latter. The factorial representation of the HJM framework has been precisely studied in Benth and Koekebakker [16] to model the electricity forward curve, giving constraints in the volatility terms to ensure no arbitrage. Koekebakker and Ollmar [58] perform a Principal Component Analysis to point out that two factors can explain 75% of the electricity forward contracts in the Norwegian market, and more than 10 factors are needed to explain 95%. They argue that, due to the non-storability of electricity, there is a weak correlation between short-term and long-term events. In Keppo *et al.* [56], a one-factor model is designed for each maturity date, having correlations between the Brownian motions for distinct dates. In Kiesel *et al.* [57], a two-factor model is described, with a specification of the volatility terms allowing to reproduce the classic behaviour of prices, especially the empirical evidence of the Samuelson effect (the volatility of prices increases as time to maturity decreases) and to ensure non-zero volatility for long-term forward prices.

In this chapter, we adopt this two-factor modeling with a parameter driving the Samuelson effect and stochastic volatility processes, that are left totally unspecified. We propose a way of efficiently estimate the scalar parameter in our semi-parametric framework in which stochastic volatility processes are first considered as nuisance elements. Once this parameter is estimated, we propose a point estimation of the volatility functions.

2.1.2 Setting

On some filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, we consider a d-dimensional Itô semi-martingale $X = (X_t)_{t\geq 0}$ with components X^j, for $j = 1, \dots, d$, of the form

$$X_t^j = X_0^j + \int_0^t b_s^j ds + \int_0^t e^{-\vartheta(T_j - s)} \sigma_s dB_s + \int_0^t \overline{\sigma}_s d\overline{B}_s, \tag{2.1}$$

where $X_0^j \in \mathbb{R}$ is an initial condition, $B = (B_t)_{t\geq 0}$ and $\overline{B} = (\overline{B}_t)_{t\geq 0}$ are two independent Brownian motions, ϑ and T_j are positive numbers and $\sigma = (\sigma_t)_{t\geq 0}$, $\overline{\sigma} = (\overline{\sigma}_t)_{t\geq 0}$, $b^j = (b_t^j)_{t\geq 0}$ are càdlàg adapted processes.

We assume that for some $T > 0$, we have

$$T \leq T_1 < \ldots < T_d$$

and that the T_i are known. Moreover, we observe X at times

$$0, \Delta_n, 2\Delta_n, \dots, n\Delta_n = T$$

Asymptotics are taken as $n \to \infty$, so we work in a standard high-frequency framework. In this setting, it is impossible to identify the components b^i, so we are left with trying to estimate the parameter ϑ and the random components $t \rightsquigarrow \sigma_t$ (or rather σ_t^2) and $t \rightsquigarrow \overline{\sigma}_t$ (or $\overline{\sigma}_t^2$) over the time interval $[0, T]$ with the best possible rate of convergence. This is not always possible and will require regularity assumptions.

2.1.3 Main results and organization of the chapter

In Section 2.2.1, we provide an estimator of ϑ, based on quadratic variation, in the above observation scheme. We will explain that while we cannot perform estimation when the number of observed processes d is equal to 1, the case $d = 2$ is statistically regular, and by approaching the quadratic variation of X^1, X^2 and $X^2 - X^1$, we derive an estimator $\hat{\vartheta}_{2,n}$ of ϑ, which is $\Delta_n^{-1/2}$-consistent. We shall make the assumption that σ and $\overline{\sigma}$ are positive processes. Using the theory of statistics for diffusion processes and relying on the tools of stable convergence in law, which are for instance summarized in [49, 62], we show that

$$\Delta_n^{-1/2}(\hat{\vartheta}_{2,n} - \vartheta) \to \mathcal{N}(0, V_\vartheta(\sigma, \overline{\sigma})),$$

stably in law, where $\mathcal{N}(0, V_\vartheta(\sigma, \overline{\sigma}))$ is a random variable defined on an extension of the original space $(\Omega, \mathcal{F}, \mathbb{P})$, and which, conditionally to $\mathcal{F}$, is Gaussian, centered, with variance $V_\vartheta(\sigma, \overline{\sigma})$.
When $d \geq 3$ processes are observed, the model is somehow degenerate, as it had been reported by Jeffrey *et al.* [55] in a similar context, because the d processes are driven by less than d Brownian motions. The remaining source of randomness is the drift process, and while we shall find a Δ_n^{-1}-consistent estimator $\hat{\vartheta}_{3,n}$ for ϑ, we will need that b has some Besov regularity in expectation, as will be made precise by Assumption 2.1, to establish a limit theorem stating that $\Delta_n^{-1}(\hat{\vartheta}_{3,n} - \vartheta)$ converges in probability to some $\mathcal{F}$-measurable random variable.
All the results for $d = 2, 3$ processes will be stated in Theorem 2.1.

In Section 2.2.2, we lead a classic nonparametric estimation procedure to get point estimates of σ_t^2 and $\overline{\sigma}_t^2$ when $d = 2$, which is yet not an usual nonparametric problem, as

1. σ and $\overline{\sigma}$ are random themselves, so that we do not estimate them pointwise, instead we estimate pointwise the trajectories $\big(\sigma_t^2(\omega)\big)_t$, $\big(\overline{\sigma}_t^2(\omega)\big)_t$, which are realizations of the volatility processes;
2. an increment $\Delta_i^n X$ is the sum of two stochastic integrals, in which the volatility processes have different regularities.

We have to separate, in some way, the parts of the random increments that are linked to each of the Brownian integrals, to be able to get estimates of each process. We shall then derive estimators $\widehat{\sigma}_n^2$ and $\widehat{\overline{\sigma}}_n^2$ of σ^2 and $\overline{\sigma}^2$ and in Theorem 2.2, adding Assumption 2.2 stating that the volatility processes are Hölder in expectation, it will be shown that each of those point estimators is $\Delta_n^{-\alpha/(2\alpha+1)}$-consistent, where α is the lowest of two values of the Hölder regularities of σ^2 and $\overline{\sigma}^2$.

In Section 2.2.3, referring to the theory of semiparametric estimation, reported for instance in the 25$^{\text{th}}$ chapter of [73], we compute a lower bound $V_\vartheta^{\text{opt}}(\sigma, \overline{\sigma})$ for the limit variance while estimating ϑ with $d = 2$ observed processes, for deterministic volatility functions, in Theorem 2.3. As soon as $\overline{\sigma}$ is not constant, this bound is lower than $V_\vartheta(\sigma, \overline{\sigma})$. Then, we derive an estimator $\tilde{\vartheta}_{2,n}$ such that

$$\Delta_n^{-1/2}(\tilde{\vartheta}_{2,n} - \vartheta) \to \mathcal{N}(0, V_\vartheta^{\text{opt}}(\sigma, \overline{\sigma}))$$

stably in law, where $\mathcal{N}(0, V_\vartheta^{\mathrm{opt}}(\sigma,\overline{\sigma}))$ is a random variable defined on an extension of the original probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and which, conditionally to $\mathcal{F}$, is Gaussian, centered, with variance $V_\vartheta^{\mathrm{opt}}(\sigma,\overline{\sigma})$. This estimator is efficient in the sense that after conditioning on $\mathcal{F}$, the limit law has the lower bound $V_\vartheta^{\mathrm{opt}}(\sigma,\overline{\sigma})$ as its own variance. This is stated in Theorem 2.4.

We shall then state Proposition 2.1, stating that we have a Δ_n^{-1}-consistent estimator when $d > 3$ processes are available. We are not yet able to give conditions under which it is more suitable to use $d > 3$ processes instead of just $d = 3$.

We perform some numerical experiments in Section 2.3, using both simulated and real data from the electricity forward markets in order to compare the behaviours of the estimators in various configurations, and the proofs of the theorems are in Section 2.4.

2.2 Construction of the estimators and convergence results

2.2.1 Rate-optimal estimation of ϑ

The case $d = 1$

In that setting, it is impossible to identify ϑ from data $X_{i\Delta_n}, i = 1, \dots, n$ asymptotically when $t \rightsquigarrow \sigma_t$ and $t \rightsquigarrow \overline{\sigma}_t$ are unknown. Indeed X has the same law under the choice of $(\vartheta, \sigma, \overline{\sigma})$ and $(\vartheta + 1, e^{T_1 - \cdot}\sigma, \overline{\sigma})$.

The case $d = 2$

This is the statistically most regular case. Set, as usual $\Delta_i^n X = X_{i\Delta_n} - X_{(i-1)\Delta_n}$ (componentwise). From the convergences

$$\sum_{i=1}^n (\Delta_i^n X^j)^2 \to \int_0^T \big(e^{-2\vartheta(T_j - t)}\sigma_t^2 + \overline{\sigma}_t^2\big)dt, \quad j = 1, 2$$

and

$$\sum_{i=1}^n (\Delta_i^n X^2 - \Delta_i^n X^1)^2 \to \int_0^T (e^{-\vartheta T_2} - e^{-\vartheta T_1})^2 e^{2\vartheta t}\sigma_t^2 dt$$

in probability, we also obtain the convergence of the ratio

$$\Psi_{T_1,T_2}^n = \frac{\sum_{i=1}^n (\Delta_i^n X^2 - \Delta_i^n X^1)^2}{\sum_{i=1}^n \big((\Delta_i^n X^2)^2 - (\Delta_i^n X^1)^2\big)} \to \frac{\big(e^{-\vartheta T_2} - e^{-\vartheta T_1}\big)^2}{e^{-2\vartheta T_2} - e^{-2\vartheta T_1}} = \psi_{T_1,T_2}(\vartheta),$$

in probability. The function $\vartheta \rightsquigarrow \psi_{T_1,T_2}(\vartheta)$ maps $(0, \infty)$ onto $(-1, 0)$ and this leads to a first $\Delta_n^{-1/2}$-consistent estimation strategy by setting

$$\hat{\vartheta}_{2,n} = \psi_{T_1,T_2}^{-1}\big(\Psi_{T_1,T_2}^n\big)$$

whenever $\Psi_{T_1,T_2}^n \in (-1, 0)$ and 0 otherwise.

The case $d = 3$

Since X is driven by two Brownian motions, the underlying statistical model becomes degenerate. Indeed, assume first that $b^1 = b^2 = b^3$. Then, we readily obtain

$$\frac{\Delta_i^n X^2 - \Delta_i^n X^1}{\Delta_i^n X^3 - \Delta_i^n X^2} = \frac{e^{-\vartheta T_2} - e^{-\vartheta T_1}}{e^{-\vartheta T_3} - e^{-\vartheta T_2}}$$

which is invertible as a function of ϑ. It is thus possible to identify ϑ exactly from the observation of a single increment of X! When the b^j are not all equal, the situation is still somehow degenerate, as we can eliminate all volatility components by taking linear combinations of the observed increments. The lowest-order remaining term is the drift process, so that we could expect to find Δ_n^{-1}-consistent estimators instead of $\Delta_n^{-1/2}$-consistent ones. We then have

$$\Psi^n_{T_1,T_2,T_3} = \frac{\sum_{i=1}^n (\Delta_i^n X^3 - \Delta_i^n X^2)^2}{\sum_{i=1}^n (\Delta_i^n X^2 - \Delta_i^n X^1)^2} \to \Big(\frac{e^{-\vartheta T_3} - e^{-\vartheta T_2}}{e^{-\vartheta T_2} - e^{-\vartheta T_1}}\Big)^2 = \psi_{T_1,T_2,T_3}(\vartheta),$$

say. The function $\vartheta \rightsquigarrow \psi_{T_1,T_2,T_3}(\vartheta)$ maps $(0,\infty)$ onto $\big(0, \big(\frac{T_3-T_2}{T_2-T_1}\big)^2\big)$ and is also invertible (see Lemma 2.4.2), leading to the estimator

$$\hat{\vartheta}_{3,n} = \psi^{-1}_{T_1,T_2,T_3}\big(\Psi^n_{T_1,T_2,T_3}\big)$$

whenever $\Psi^n_{T_1,T_2,T_3} \in \big(0, \big(\frac{T_3-T_2}{T_2-T_1}\big)^2\big)$ and 0 otherwise.

Convergence results

We need some assumption about the regularity of the processes b, σ and $\overline{\sigma}$. For a random process $X = (X_t)_{0\le t\le T}$, introduce the following modulus of continuity:

$$\omega(X)_t = \sup_{|h|\le t} \Big(\int_0^T \mathbb{E}\big[(X_{s+h} - X_s)^2\big]\,ds\Big)^{1/2}.$$

Assumption 2.1. *The processes σ and $\overline{\sigma}$ are almost surely positive. Moreover, for some $s > 1/2$, we have $\sup_{t\in[0,T]} t^{-s}\omega(b^j)_t < \infty$ for every $j = 1, \dots, d$.*

To state the convergence results, we need some notation. Set

$$\overline{b}_t = 2(e^{-\vartheta T_2} - e^{-\vartheta T_1})(e^{-\vartheta T_3} - e^{-\vartheta T_2})\big((e^{-\vartheta T_2} - e^{-\vartheta T_1})(b_t^3 - b_t^2) - (e^{-\vartheta T_3} - e^{-\vartheta T_2})(b_t^2 - b_t^1)\big)$$

and

$$\widetilde{b}_T = (e^{-\vartheta T_2} - e^{-\vartheta T_1})^2 \int_0^T (b_t^3 - b_t^2)^2 dt - (e^{-\vartheta T_3} - e^{-\vartheta T_2})^2 \int_0^T (b_t^2 - b_t^1)^2 dt.$$

We also set

$$D_3 = (e^{-\vartheta T_3} - e^{-\vartheta T_2})\big((e^{-\vartheta T_3} - e^{-\vartheta T_2})(T_2 e^{-\vartheta T_2} - T_1 e^{-\vartheta T_1}) - (e^{-\vartheta T_2} - e^{-\vartheta T_1})(T_3 e^{-\vartheta T_3} - T_2 e^{-\vartheta T_2})\big).$$

Theorem 2.1. *Work under Assumption 2.1.*

1. *For the case $d = 2$, we have*
$$\Delta_n^{-1/2}(\hat{\vartheta}_{2,n} - \vartheta) \to \mathcal{N}\big(0, V_\vartheta(\sigma, \overline{\sigma})\big)$$
in distribution as $n \to \infty$, where $\mathcal{N}\big(0, V_\vartheta(\sigma, \overline{\sigma})\big)$ is a random variable which, conditionally to $\mathcal{F}$, is centered normal with variance
$$V_\vartheta(\sigma, \overline{\sigma}) = \frac{1}{(T_2 - T_1)^2}(e^{\vartheta T_2} - e^{\vartheta T_1})^2 \frac{\int_0^T e^{2\vartheta t}\sigma_t^2 \overline{\sigma}_t^2 dt}{\big(\int_0^T e^{2\vartheta t}\sigma_t^2 dt\big)^2}.$$

2. *For the case $d = 3$ we have*
$$\Delta_n^{-1}(\hat{\vartheta}_{3,n} - \vartheta) \to \frac{\widetilde{b}_T + \int_0^T \bar{b}_t e^{\vartheta t}\sigma_t dB_t}{2(e^{-\vartheta T_2} - e^{-\vartheta T_1})D_3 \int_0^T e^{2\vartheta t}\sigma_t^2 dt}$$
in probability as $n \to \infty$.

2.2.2 Rate-optimal estimation of the volatility processes

Construction of an estimator

We start with the classic observation that for any sufficiently regular test function $g : [0, T] \to \mathbb{R}$, we have, for any $j = 1, \ldots, d$,
$$\sum_{i=1}^n g((i-1)\Delta_n)\big(\Delta_i^n X^j\big)^2 \to \int_0^T g(s)d\langle X^j\rangle_s = \int_0^T g(s)\big(e^{-2\vartheta(T_j - s)}\sigma_s^2 + \overline{\sigma}_s^2\big)ds \tag{2.2}$$
in probability as $n \to \infty$. Therefore, picking a function g that mimics the Dirac mass $\delta_t(ds)$ at point t, we can asymptotically identify
$$e^{-2\vartheta(T_1 - t)}\sigma_t^2 + \overline{\sigma}_t^2 \ \text{ and } \ e^{-2\vartheta(T_2 - t)}\sigma_t^2 + \overline{\sigma}_t^2$$
by applying (2.2) for $j = 1, 2$. We thus identify σ_t^2 and $\overline{\sigma}_t^2$ as well by inverting a 2×2 linear system, namely
$$\begin{pmatrix} \sigma_t^2 \\ \overline{\sigma}_t^2 \end{pmatrix} = \mathcal{M}(\vartheta)_t \begin{pmatrix} e^{-2\vartheta(T_1 - t)}\sigma_t^2 + \overline{\sigma}_t^2 \\ e^{-2\vartheta(T_2 - t)}\sigma_t^2 + \overline{\sigma}_t^2 \end{pmatrix}$$
where
$$\mathcal{M}(\vartheta)_t = \frac{1}{e^{-2\vartheta(T_1 - t)} - e^{-2\vartheta(T_2 - t)}} \begin{pmatrix} 1 & -1 \\ -e^{-2\vartheta(T_2 - t)} & e^{-2\vartheta(T_1 - t)} \end{pmatrix}.$$
For $\varpi_n > 0$ and $h_n > 0$, define the estimators
$$\begin{pmatrix} \widehat{\sigma}_{n,t}^2 \\ \widehat{\overline{\sigma}}_{n,t}^2 \end{pmatrix} = h_n^{-1}\mathcal{M}(\max\{\hat{\vartheta}_{2,n}, \varpi_n\})_t \sum_{t - h_n \le (i-1)/n < t} \begin{pmatrix} (\Delta_i^n X^1)^2 \\ (\Delta_i^n X^2)^2 \end{pmatrix}. \tag{2.3}$$

As in classic nonparametric estimation, we need the bandwidth h_n to realize the usual compromise between bias and variance. Besides, we introduce the sequence ϖ_n so that we never plug a null value of $\hat{\vartheta}_{2,n}$ into $\mathcal{M}(\vartheta)_t$, as it is not defined.

Convergence results

We need an additional regularity assumption on the volatility processes σ and $\overline{\sigma}$.

Assumption 2.2. *There exists a constant $c > 0$ and $\alpha \geq 1/2$ such that for every $t, s \in [0, T]$, we have*

$$\mathbb{E}\big[|\sigma_t^2 - \sigma_s^2|^2\big] + \mathbb{E}\big[|\overline{\sigma}_t^2 - \overline{\sigma}_s^2|^2\big] \leq c|t-s|^{2\alpha}. \tag{2.4}$$

Theorem 2.2. *Work under Assumptions 2.1 and 2.2. Let h_n be specified by*

$$h_n = \Delta_n^{1/(2\alpha+1)},$$

and let ϖ_n be any sequence of positive numbers that decreases to 0.
Then the sequence

$$\Delta_n^{-\alpha/(2\alpha+1)} \sup_{t \in [h_n, T]} \Big[\big|\widehat{\sigma}_{n,t}^2 - \sigma_t^2\big| + \big|\widehat{\overline{\sigma}}_{n,t}^2 - \overline{\sigma}_t^2\big| \Big]$$

is tight. This implies that

$$\Delta_n^{-\alpha/(2\alpha+1)} \Big[\big|\widehat{\sigma}_{n,t}^2 - \sigma_t^2\big| + \big|\widehat{\overline{\sigma}}_{n,t}^2 - \overline{\sigma}_t^2\big| \Big]$$

is tight, uniformly for $t \in \mathcal{D}$, where $\mathcal{D}$ is any compact included in $(0, T]$.

2.2.3 Efficient estimation of ϑ when $d = 2$

We look for the best attainable variance among rate-optimal estimators of ϑ that are asymptotically normal. However, we do not have a statistical model in the classic sense, in which the parameter would simply be $(\vartheta, \sigma, \overline{\sigma})$, because of the nuisance parameters σ and $\overline{\sigma}$ which are random processes themselves! In order to bypass this difficulty, we first restrict our attention to the case where σ and $\overline{\sigma}$ are deterministic functions, which enables us to identify our data within a semiparametric regular statistical model. Thanks to classic bounds on semiparametric estimation, we can explicitly compute the optimal (best achievable) variance $V_\vartheta^{\mathrm{opt}}(\sigma, \overline{\sigma})$. In a second step, allowing σ and $\overline{\sigma}$ to be random again, we build a one-step correction of our preliminary estimator $\hat{\vartheta}_{2,n}$ which has the property of being asymptotically mixed normal, with (conditional) variance equal to $V_\vartheta^{\mathrm{opt}}(\sigma, \overline{\sigma})$, *i.e.* thus achieving the optimal variance along deterministic paths.

Lower bounds

Consider the statistical experiment $\mathcal{E}^n$ generated by data $(\Delta_i^n X^1, \Delta_i^n X^2, i = 1, \ldots, n)$ with

$$X_t^i = X_0^i + \int_0^t e^{-\vartheta(T_i - s)} \sigma_s dB_s + \int_0^t \overline{\sigma}_s d\overline{B}_s, \quad i = 1, 2, \tag{2.5}$$

with parameter $(\vartheta, \sigma, \overline{\sigma}) \in \Theta \times \Sigma(c, \tilde{c})$, with $\Theta = (0, \infty)$ and $\Sigma(c, \tilde{c})$ being the space of positive (deterministic) functions $(\sigma, \overline{\sigma})$ defined on $[0, T]$, satisfying (2.4) of Assumption 2.2 with constant c, being moreover bounded below by some $\tilde{c} > 0$.

Theorem 2.3. *Let $\hat{\vartheta}_n$ be an estimator of ϑ in the experiment $\mathcal{E}^n$ such that $\Delta_n^{-1/2}(\hat{\vartheta}_n - \vartheta)$ converges to $\mathcal{N}\big(0, V_\vartheta(\sigma, \overline{\sigma})\big)$ in distribution as $n \to \infty$. Then*

$$V_\vartheta(\sigma,\overline{\sigma}) \geq V_\vartheta^{\mathrm{opt}}(\sigma,\overline{\sigma}) = \frac{1}{(T_2-T_1)^2}(e^{\vartheta T_2} - e^{\vartheta T_1})^2 \Big(\int_0^T \frac{e^{2\vartheta t}\sigma_t^2}{\overline{\sigma}_t^2} dt \Big)^{-1}.$$

Construction of an efficient procedure

This is the most delicate part of this chapter. By representation (2.5), we see that the $(\Delta_i^n X^1, \Delta_i^n X^2)$ are independent for $i = 1, \dots, n$. Moreover, $(\Delta_i^n X^1, \Delta_i^n X^2)$ is a centered Gaussian vector with explicit covariance structure

$$\mathbb{E}\big[(\Delta_i^n X^1)^2\big] = \int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)}\sigma_t^2 dt + \int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t^2 dt,$$

$$\mathbb{E}\big[(\Delta_i^n X^2)^2\big] = \int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_2-t)}\sigma_t^2 dt + \int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t^2 dt$$

and

$$\mathbb{E}\big[\Delta_i^n X^1 \Delta_i^n X^2\big] = \int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1+T_2-2t)}\sigma_t^2 dt + \int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t^2 dt.$$

Let us further denote by $f_{\vartheta,\sigma,\overline{\sigma}}$ its density function w.r.t. the Lebesgue measure on $\mathbb{R}^2$. If the nuisance parameters $(\sigma,\overline{\sigma})$ were known, then an optimal (efficient) procedure could be obtained by a one-step correction of the type

$$\hat{\vartheta}_n = \hat{\vartheta}_{2,n} + \frac{\sum_{i=1}^n \ell^i_{\vartheta,\sigma,\overline{\sigma}}(\hat{\vartheta}_{2,n})}{\sum_{i=1}^n \big(\ell^i_{\vartheta,\sigma,\overline{\sigma}}(\hat{\vartheta}_{2,n})\big)^2}$$

where $\ell^i_{\vartheta,\sigma,\overline{\sigma}}(\hat{\vartheta}_{2,n}) = \partial_\vartheta \log f_{\vartheta,\sigma,\overline{\sigma}}(\Delta_i^n X^1, \Delta_i^n X^2)$ is the score function associated to the observation $(\Delta_i^n X^1, \Delta_i^n X^2)$, see for instance Section 8.9 in [73]. However, this oracle procedure is not achievable and we need to invoke the theory of semiparametric efficiency (see for instance the 25th chapter of [73]). In the presence of an extra nuisance parameter $(\sigma,\overline{\sigma})$, we consider instead the so-called efficient score

$$\widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}}(\vartheta) = \ell^i_{\vartheta,\sigma,\overline{\sigma}}(\vartheta) - \Pi \ell^i_{\vartheta,\sigma,\overline{\sigma}}(\vartheta),$$

where Π is the projection operator onto the tangent space associated to a one-dimensional perturbation around the true (unknown) value $(\sigma,\overline{\sigma})$. It turns out that we indeed have a simple and explicit formula for $\widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}}(\vartheta)$ which enables us to derive a one-step correction formula using $\widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}}(\vartheta)$ and plug-in estimators in order to achieve the optimal bound.

For technical reason, we replace from now on $\hat{\vartheta}_{2,n}$ by $\Delta_n^{1/2}\lfloor \Delta_n^{-1/2}\hat{\vartheta}_{2,n}\rfloor$ and we still write $\hat{\vartheta}_{2,n}$ for simplicity. Likewise, we implicitly replace the estimators $\widehat{\overline{\sigma}}^2_{n,t}$ defined in (2.3) by

$\max\{\widehat{\overline{\sigma}}_{n,t}^2, \tilde{c}^2\}$, where $\tilde{c}$ is the lower bound associated to $\Sigma(c, \tilde{c})$ in the definition of the experiment $\mathcal{E}^n$. For $i = 1, \ldots, n$, define

$$\widetilde{\ell}_{\overline{\sigma}}(\vartheta)^i = \frac{(\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta(T_2 - T_1)}\Delta_i^n X^1)e^{-\vartheta(T_2-T_1)}(T_2 - T_1)}{(1 - e^{-\vartheta(T_2-T_1)})^3 \int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t^2 dt}.$$

Theorem 2.4. *Work under Assumptions 2.1 and 2.2 with $\alpha > 1/2$. For $i = 1, \ldots, n$, the efficient score for the parameter ϑ associated to $(\Delta_i^n X^1, \Delta_i^n X^2)$ in the experiment $\mathcal{E}^n$ is given by $\widetilde{\ell}_{\overline{\sigma}}(\vartheta)^i$. Moreover, the estimator $\tilde{\vartheta}_{2,n}$ defined by*

$$\tilde{\vartheta}_{2,n} = \hat{\vartheta}_{2,n} + \frac{\sum_{i \in \mathcal{I}_n} \widetilde{\ell}\big(\hat{\vartheta}_{2,n}, \widehat{\overline{\sigma}}_n^2\big)^i}{\sum_{i \in \mathcal{I}_n} \Big(\widetilde{\ell}\big(\hat{\vartheta}_{2,n}, \widehat{\overline{\sigma}}_n^2\big)^i\Big)^2}$$

with $\mathcal{I}_n = \{i = 1, \ldots, n | h_n \leq (i-1)\Delta_n < T\}$ and

$$\widetilde{\ell}\big(\hat{\vartheta}_{2,n}, \widehat{\overline{\sigma}}_n^2\big)^i = \frac{(\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\hat{\vartheta}_{2,n}(T_2-T_1)}\Delta_i^n X^1)e^{-\hat{\vartheta}_{2,n}(T_2-T_1)}(T_2 - T_1)}{(1 - e^{-\hat{\vartheta}_{2,n}(T_2-T_1)})^3 \Delta_n \widehat{\overline{\sigma}}_{n,(i-1)\Delta_n}^2}$$

satisfies

$$\Delta_n^{-1/2}\big(\tilde{\vartheta}_{2,n} - \vartheta\big) \to \mathcal{N}\big(0, V_\vartheta^{\mathrm{opt}}(\sigma, \overline{\sigma})\big)$$

in distribution as $n \to \infty$. Moreover, the result is still valid if σ and $\overline{\sigma}$ are random processes such that $\mathbb{P}\big((\sigma, \overline{\sigma}) \in \Sigma(c, \tilde{c})\big) = 1$. In that case, the limiting distribution is, conditionally to $\mathcal{F}$, centered Gaussian with (conditional) variance $V_\vartheta^{\mathrm{opt}}(\sigma, \overline{\sigma})$.

This result shows that the lower bound $V_\vartheta^{\mathrm{opt}}(\sigma, \overline{\sigma})$ can be attained, and therefore that efficient estimation can be performed (which has a sense only for deterministic volatility functions). Using Cauchy-Schwarz inequality, is is easy to prove that the expression of the limit variance is equal to the one we got in Theorem 2.1 for $\Delta_n^{-1/2}(\hat{\vartheta}_{2,n} - \vartheta)$ if and only if $\overline{\sigma}$ is constant over the interval $[0, T]$. Otherwise, efficient estimation is more accurate than the one in the first part of Theorem 2.1.

Remark 2.1. *Our current proof method does not allow us to extend Theorem 2.4 to the limit case $\alpha = 1/2$. To do so, other technical tools would be required. Whether the theorem is valid or not when $\alpha = 1/2$ is still an open problem.*

2.2.4 Discussion on the case $d \geq 3$

In Section 2.2.1, we built estimators of ϑ for $d = 2$ and $d = 3$, emphasizing the differences between those two cases. When $d > 3$, we meet the same problem of degeneracy as when $d = 3$: the d processes are driven by 2 Brownian motions only. We may therefore build an estimator similar to the one with three processes. We have

$$\Psi_{T_{1..d}}^n = \sum_{j=3}^d \frac{\sum_{i=1}^n (\Delta_i^n X^j - \Delta_i^n X^{j-1})^2}{\sum_{i=1}^n (\Delta_i^n X^2 - \Delta_i^n X^1)^2} \to \sum_{j=3}^d \Big(\frac{e^{-\vartheta T_j} - e^{-\vartheta T_{j-1}}}{e^{-\vartheta T_2} - e^{-\vartheta T_1}}\Big)^2 = \psi_{T_{1..d}}(\vartheta).$$

The function $\vartheta \rightsquigarrow \psi_{T_{1..d}}(\vartheta)$ maps $(0,\infty)$ onto $\big(0, \sum_{j=3}^d \big(\frac{T_j-T_{j-1}}{T_2-T_1}\big)^2\big)$ and is invertible as the sum of $d-2$ monotone functions (see Lemma 2.4.2). We can thus propose the estimator

$$\hat{\vartheta}_{d,n} = \psi_{T_{1..d}}^{-1}\big(\Psi_{T_{1..d}}^n\big)$$

whenever $\Psi_{T_{1..d}}^n \in \big(0, \sum_{j=3}^d \big(\frac{T_j-T_{j-1}}{T_2-T_1}\big)^2\big)$ and 0 otherwise.

It is possible to state the following proposition, using the notation

$$\begin{aligned}
\overline{b}_t^d =& 2(e^{-\vartheta T_2} - e^{-\vartheta T_1}) \sum_{j=3}^d (e^{-\vartheta T_j} - e^{-\vartheta T_{j-1}}) \big[(e^{-\vartheta T_2} - e^{-\vartheta T_1})(b_t^j - b_t^{j-1}) \\
&\qquad\qquad\qquad\qquad\qquad\qquad - (e^{-\vartheta T_j} - e^{-\vartheta T_{j-1}})(b_t^2 - b_t^1)\big], \\
\widetilde{b}_T^d =& (e^{-\vartheta T_2} - e^{-\vartheta T_1})^2 \int_0^T \sum_{j=3}^d (b_t^j - b_t^{j-1})^2 dt - \sum_{j=3}^d (e^{-\vartheta T_j} - e^{-\vartheta T_{j-1}})^2 \int_0^T (b_t^2 - b_t^1)^2 dt, \\
D_d =& \sum_{j=3}^d (e^{-\vartheta T_j} - e^{-\vartheta T_{j-1}}) \big[(e^{-\vartheta T_j} - e^{-\vartheta T_{j-1}})(T_2 e^{-\vartheta T_2} - T_1 e^{-\vartheta T_1}) \\
&\qquad\qquad\qquad - (e^{-\vartheta T_2} - e^{-\vartheta T_1})(T_j e^{-\vartheta T_j} - T_{j-1} e^{-\vartheta T_{j-1}})\big].
\end{aligned}$$

Proposition 2.1. *Work under Assumption 2.1. We have*

$$\Delta_n^{-1}(\hat{\vartheta}_{d,n} - \vartheta) \to \frac{\widetilde{b}_T^d + \int_0^T \overline{b}_t^d e^{\vartheta t} \sigma_t dB_t}{2(e^{-\vartheta T_2} - e^{-\vartheta T_1}) D_d \int_0^T e^{2\vartheta t} \sigma_t^2 dt}$$

in probability as $n \to \infty$.

We will prove this proposition in a more general context, while stating Theorem 3.2 in Chapter 3.

A natural question arises while defining this new estimator: are we able to determine if using $d > 3$ processes is better than using $d = 3$ processes only? As the convergence rate is the same, the criterion should be the comparison of the limits in probability of $\Delta_n^{-1}(\hat{\vartheta}_{3,n} - \vartheta)$ and $\Delta_n^{-1}(\hat{\vartheta}_{d,n} - \vartheta)$. So far, we did not manage to find sufficient conditions so that one of those limits is closest to zero than the other one is. This is an open problem, and in numerical experiments we shall compute all the estimators that are available in order to compare them.

2.3 Numerical implementation

In this section we give illustration results of the proposed estimation procedures. We start this section by briefly presenting the context of forward contracts observed in the electricity markets, which motivates our mathematical setting. Then the estimation results are shown both on simulated data and real observations.

2.3.1 Context of electricity forward contracts

The prices of existing forward contracts in the electricity markets are characterized by three time components: the quotation date t and the dates T_s and T_e of respectively starting and ending power delivery. Therefore, a forward contract $F(t, T_s, T_e)$ will deliver to the holder 1 MWh of electricity, in a continuous way between dates T_s and T_e. Such a contract may be bought during a quotation period $[t_0, T]$ with $T < T_s$ and it is no more available once $t > T$. The classic observed contracts are of various delivery periods: one week, one month, one quarter (three months), one season (6 months) or one year. Table 2.1 shows an example of available forward contracts in the French Market on May 23$^{\text{rd}}$, 2015. For example, the contract called "June 2015" will deliver to the holder 1 MWh of electricity for all the hours between June 1$^{\text{st}}$ (this is T_s) and June 30$^{\text{th}}$ of 2015 (T_e). This table also introduces the contracts of relative maturity (denoted by the "ahead" formulation). A "ahead" contract is a contract with constant delivery period but with changing delivery dates. For example, the 2-month-ahead contract is the forward contract "July 2015" when it is quoted on May 31$^{\text{th}}$ of 2015 (2 months ahead from the quotation date), and becomes the forward contract "August 2015" on June 1$^{\text{st}}$, 2015 (a jump of contract to stay 2 months ahead from the quotation date).

In this study we will only consider the 6 observable monthly contracts (*i.e.* $T_e - T_s = 1$ month) to estimate ϑ and the volatility processes σ and $\overline{\sigma}$. Also, for simplicity, we will drop T_e from the notation. In the context of simulated data, we will simulate prices of $F(t, T_s) = F(t, T_s, T_e)$, the forward delivering 1 MWh during the period $[T_s, T_e]$. In the context of real data, the price $F(t, T_s)$ is observable.

Product	Example: May 23$^{\text{rd}}$, 2015		
	Name of the product	Begin of delivery	End of delivery
1 Month Ahead	June 2015	2015-06-01	2015-06-30
2 Month Ahead	July 2015	2015-07-01	2015-07-31
3 Month Ahead	August 2015	2015-08-01	2015-08-31
4 Month Ahead	September 2015	2015-09-01	2015-09-30
5 Month Ahead	October 2015	2015-10-01	2015-10-31
6 Month Ahead	November 2015	2015-11-01	2015-11-30
1 Quarter Ahead	3$^{\text{rd}}$ quarter 2015	2015-07-01	2015-09-30
2 Quarter Ahead	4$^{\text{th}}$ quarter 2015	2015-10-01	2015-12-31
3 Quarter Ahead	1$^{\text{st}}$ quarter 2016	2016-01-01	2016-03-31
1 Year Ahead	Year 2016	2016-01-01	2016-12-31

Table 2.1 – Data available each day

2.3.2 Results on simulated data

The objective of this section is to study the estimator's behaviour on a simulated data set, where the log-prices of the forward contracts are simulated according to the two-factor model described in (2.1). The parameter values are chosen to be close to values estimated on real data: in [57], the volatility processes are constant, and the estimated values are $\sigma = 0.37\ \mathrm{y}^{-1/2}$ and $\overline{\sigma} = 0.15\ \mathrm{y}^{-1/2}$. Here we use a CIR-like model (the Cox-Ingersoll-Ross model for interest rates has been introduced in [26], in 1985), to emphasize the fact that our model may also be used in the context of interest rates modeling (this is indeed where it comes from, see [42]). Our parameters are

$$b_t^j = 3.65 \cdot 10^{-1}(\log(30) - X_t^j), \sigma_t = 0.37\Sigma_t^d \text{ and } \sigma_t = 0.15\Sigma_t^d,$$

with $\Sigma_t^d = \sqrt{\frac{1}{d}\sum_{j=1}^d X_t^j}$, which is the square root of the average of the d quoted log-prices. We adopt various values of ϑ (values in y^{-1}): 1.4, 10, 20, 40. The first value is the estimated parameter shown in [57] and the others are chosen to cover a wide range of possible values to observe different behaviours of our estimators. Finally, the initial value of each simulated log-price series is the logarithm of a random variable taken uniformly over the interval $[20, 40]$, which is an usual range for prices in the market of forward contracts on electricity (see also the constant 30 in the drift, in the center of that interval).
We consider different simulation configurations, all related to the situations we are facing on real data.

- 2 processes (1 month-ahead and 2 month-ahead) observed on $n = 100$ dates, with $T = T_1 = 150$ and $T_2 = 181$ days.
- 3 processes (1 month-ahead to 3 month-ahead) observed on $n = 80$ dates, with $T = T_1 = 120$, $T_2 = 150$ and $T_3 = 181$ days.
- 4 processes (1 month-ahead to 4 month-ahead) observed on $n = 60$ dates, with $T = T_1 = 90$, $T_2 = 120$, $T_3 = 151$ and $T_4 = 181$ days.
- 5 processes (1 month-ahead to 5 month-ahead) observed on $n = 40$ dates, with $T = T_1 = 59$, $T_2 = 90$, $T_3 = 120$, $T_4 = 151$ and $T_5 = 181$ days.
- 6 processes (1 month-ahead to 6 month-ahead) observed on $n = 20$ dates, with $T = T_1 = 31$, $T_2 = 59$, $T_3 = 90$, $T_4 = 120$, $T_5 = 151$ and $T_6 = 181$ days.

The decreasing number of observations corresponds to the configuration observed with real data: 2 monthly contracts are jointly observed on working days during 5 months (around 100 quotation dates) whereas 6 monthly contracts can be jointly observed only during 1 month (around 20 quotation dates). The number of observations is a bit low, as we are relying on asymptotic results. This is, of course, something that must be kept in mind in what follows. The estimation performances are now evaluated in each configuration.

In each of the configurations, we perform 100,000 simulations leading to 100,000 estimators of ϑ. Recall that we denote by $\hat{\vartheta}_{j,n}$ the estimator of ϑ from the configuration where j processes are observed, and also by $\tilde{\vartheta}_{2,n}$ the efficient estimator as described in Section 2.2.3, available in the configuration of 2 observed processes. As we said in Remark 2.1, we have not proved that the estimator $\tilde{\vartheta}_{2,n}$ is $\Delta_n^{-1/2}$-consistent and that it reaches the lower bound for the limit variance. Yet, we did not get any numerical evidence against that possibility.
Tables 2.2, 2.3, 2.4 and 2.5 give the estimation results for $\vartheta = 1.4$, 10, 20 and 40 y^{-1}, respectively. In each configuration, these tables give the number of converging instances of the

estimator and their average, and the empirical confidence interval at 95% (issued from taking the quantiles of the sample of estimated values). We must notice that some occurrences may not lead to a solution in the estimation procedure because $\Psi^n_{T_1,T_2}$ and $\Psi^n_{T_1,T_2,T_3}$, defined in Section 2.2.1, can sometimes take values outside the supports of $\psi^{-1}_{T_1,T_2}$ and $\psi^{-1}_{T_1,T_2,T_3}$.
We can see that the estimators perform quite well: except in three lines in Table 2.5, the true value of ϑ is always in the confidence interval.
Recall the result from Theorem 2.1 : the convergence toward the limit is realized at the rate $\Delta_n^{-1/2}$ for $\hat{\vartheta}_{2,n}$ and $\tilde{\vartheta}_{2,n}$, and at the rate Δ_n^{-1} for $\hat{\vartheta}_{j,n}$, $j \geq 3$. We cannot give the limit law in any case, as it is random and depends on the paths of the process X.
Finally, we may observe that adding new maturities does not improve the quality of estimation in all configurations. For instance, increasing the number of maturities may increase or decrease the length of the confidence interval, and it may shift it away from the true value of ϑ. Notice also that the one-step correction from $\hat{\vartheta}_{2,n}$ to $\tilde{\vartheta}_{2,n}$ never led to very different values.
One may refer to Section 2.5.3 in the appendices of the chapter, where some illustrative histograms of the 100,000 values of the estimators together with the true value of ϑ may be found for $\vartheta = 1.4$ and $\vartheta = 40$. In particular, it is interesting to observe the bias that occurs with $\vartheta = 40$, see Figures 2.13–2.15. That observation should prevent us from being too confident in that estimator while looking at real data.

Processes	Estimator	Instances that converged	Average	Quantile interval
2	$\hat{\vartheta}_{2,n}$	100,000	1.4216	[1.2697,1.6048]
2	$\tilde{\vartheta}_{2,n}$	100,000	1.4217	[1.2697,1.6048]
3	$\hat{\vartheta}_{3,n}$	99,962	1.3799	[0.77864,1.9250]
4	$\hat{\vartheta}_{4,n}$	100,000	1.3840	[1.0752,1.7646]
5	$\hat{\vartheta}_{5,n}$	100,000	1.3807	[1.1274,1.6864]
6	$\hat{\vartheta}_{6,n}$	100,000	1.3849	[1.0989,1.7644]

Table 2.2 – Results of the estimation on simulated data with $\vartheta = 1.4$ y^{-1}

Processes	Estimator	Instances that converged	Average	Quantile interval
2	$\hat{\vartheta}_{2,n}$	99,953	10.507	[7.2997,16.500]
2	$\tilde{\vartheta}_{2,n}$	99,953	10.507	[7.2992,16.502]
3	$\hat{\vartheta}_{3,n}$	100,000	9.9498	[9.4916,10.258]
4	$\hat{\vartheta}_{4,n}$	100,000	9.9424	[9.6307,10.195]
5	$\hat{\vartheta}_{5,n}$	100,000	9.9388	[9.6538,10.180]
6	$\hat{\vartheta}_{6,n}$	100,000	9.9511	[9.6331,10.242]

Table 2.3 – Results of the estimation on simulated data with $\vartheta = 10$ y^{-1}

Concerning the estimation results of the volatility processes σ_t^2 and $\overline{\sigma}_t^2$, we use the causal kernel $K(x) = \mathbf{1}_{(0,1](x)}$, and the bandwidth h_n for the two volatility functions is selected

Processes	Estimator	Instances that converged	Average	Quantile interval
2	$\hat{\vartheta}_{2,n}$	85,677	21.145	[10.677,47.124]
2	$\tilde{\vartheta}_{2,n}$	85,677	21.130	[10.673,47.047]
3	$\hat{\vartheta}_{3,n}$	100,000	19.672	[18.198,20.247]
4	$\hat{\vartheta}_{4,n}$	100,000	19.567	[18.233,20.204]
5	$\hat{\vartheta}_{5,n}$	100,000	19.518	[18.247,20.202]
6	$\hat{\vartheta}_{6,n}$	100,000	19.699	[18.739,20.362]

Table 2.4 – Results of the estimation on simulated data with $\vartheta = 20$ y^{-1}

Processes	Estimator	Instances that converged	Average	Quantile interval
2	$\hat{\vartheta}_{2,n}$	55,248	24.747	[10.215,56.650]
2	$\tilde{\vartheta}_{2,n}$	55,248	24.716	[10.210,56.663]
3	$\hat{\vartheta}_{3,n}$	100,000	33.904	[22.598,40.060]
4	$\hat{\vartheta}_{4,n}$	100,000	32.162	[22.204,39.689]
5	$\hat{\vartheta}_{5,n}$	100,000	31.075	[22.046,38.832]
6	$\hat{\vartheta}_{6,n}$	100,000	33.901	[26.134,39.320]

Table 2.5 – Results of the estimation on simulated data with $\vartheta = 40$ y^{-1}

by cross validation and visual inspection: as the number of data is quite poor, the empirical criterion to be minimized in the cross validation method does not always admit a minimum. We therefore retain a value of h_n near to the values that are given by cross validation when the minimization is well defined, and we check that is does not lead to obvious under- or oversmoothing. The retained value is 14 days. We also set $\varpi_n = 3.65 \cdot 10^{-2}$. In the following we show the estimators $\widehat{\sigma}_n^2$ and $\widehat{\overline{\sigma}}_n^2$ for the configuration where 2 processes are simulated on a period of 5 months (approximately 150 days), which means $T = T_1 = 150$ and $T_2 = 181$ days, with $n = 100$ dates and $\vartheta = 10$ y^{-1}. First we keep the specification $b_t^j = 3.65 \cdot 10^{-1}(\log(30) - X_t^j)$ for the drift process, but we use the constant volatility processes of [57], that is $\sigma = 0.37$ y$^{-1/2}$ and $\overline{\sigma} = 0.15$ y$^{-1/2}$. A deterministic specification allows us to compare the curve of point estimates with the deterministic function that was used to simulate the processes.
Remember that the nonparametric estimation result, Theorem 2.2, gives convergence uniformly on $[h_n, T]$. Therefore we expect that the fit is not good for values ot t being less than h_n.
We perform simulation and estimation 10,000 times, and then take the average and the quantiles of the 10,000 curves (that is, at each point t of the discretization grid, we take the average and the quantiles at 2.5% and 97.5% of the 10,000 occurrences of $\widehat{\sigma}_{n,t}^2$ and $\widehat{\overline{\sigma}}_{n,t}^2$).

Figure 2.1 gives the square of the estimated equivalent volatility function, that is the sum $e^{-2\hat{\vartheta}_{2,n}(T_1-t)}\widehat{\sigma}_{n,t}^2 + \widehat{\overline{\sigma}}_{n,t}^2$, together with the true function $e^{-2\vartheta(T_1-t)}\sigma_t^2 + \overline{\sigma}_t^2$. It shows a good estimation of this equivalent volatility, the error (between the average of the 10,000 estimators and the true value) being maximal in the two ends of the curve. The estimation of $\overline{\sigma}_t^2$, given in Figure 2.2, also performs well. However, we can observe in Figure 2.3 a bad performance of estimation of σ_t^2, especially for large values of $T - t$. This can be explained by the fact that, due to the presence of the exponential term $e^{-\vartheta(T_\cdot - t)}$, the short term factor $e^{-2\vartheta(T_\cdot - t)}\sigma_t^2$ is low when $T - t$ is large. Also, if ϑ happens to be overestimated, the estimator of σ_t^2 has to take a very high value so that the product $\sigma_t^2 e^{-2\vartheta(T_\cdot - t)}$ may fit the curve. Therefore, the estimation of σ_t should reasonably be taken into account only for small times to maturity $T - t$, where the estimation procedure seems to work well.

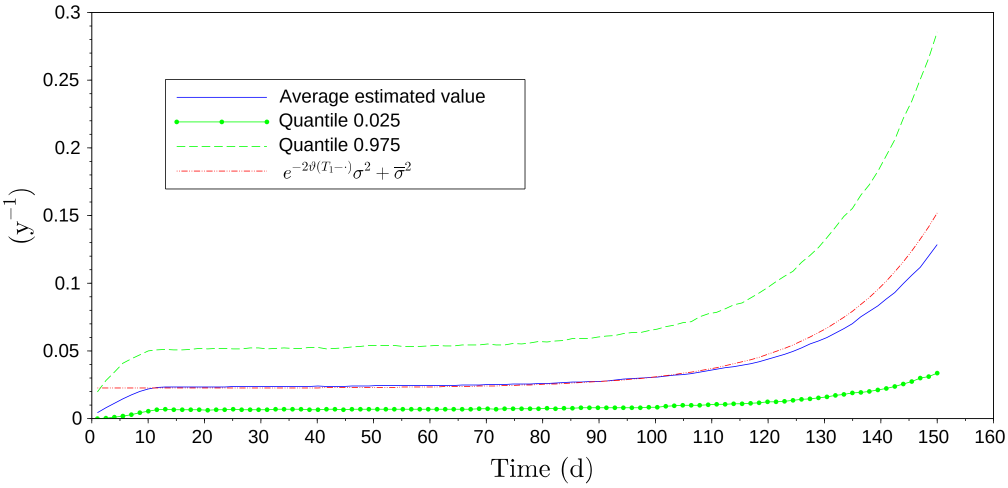

Figure 2.1 – Quantiles for the square of the equivalent volatility, with 2 processes, $\vartheta = 10$ y^{-1} and deterministic constant volatilities

Now, we are back to the specification $\sigma_t = 0.37\Sigma_t^d, \overline{\sigma}_t = 0.15\Sigma_t^d$. As the volatility processes depend on the path of X, we cannot compare visually the real volatility and its point estimators. Yet, we plot the average and the quantile curves of the 10,000 estimators for the two volatility processes and for the equivalent square volatility process, in Figures 2.4, 2.5 and 2.6. The behaviours of the series of point estimators are very similar to the ones we described while considering deterministic volatility functions.

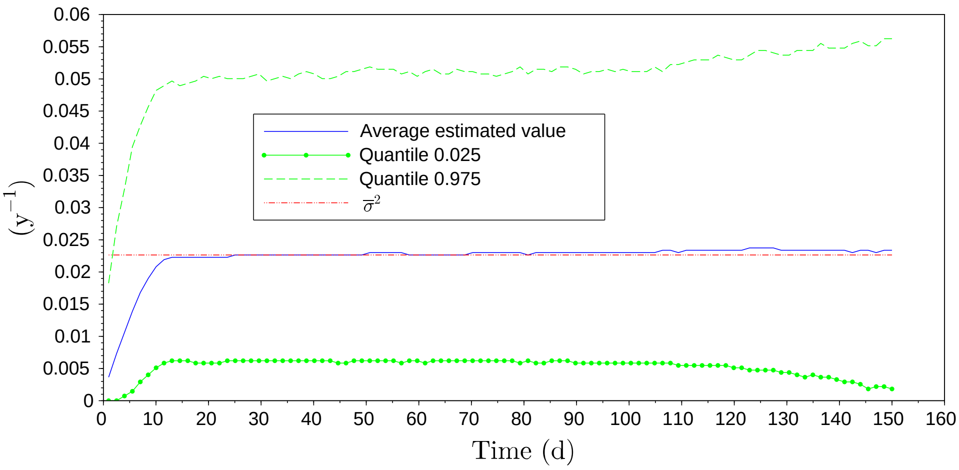

Figure 2.2 – Quantiles for the square of the long-term volatility, with 2 processes, $\vartheta = 10$ y^{-1} and deterministic constant volatilities

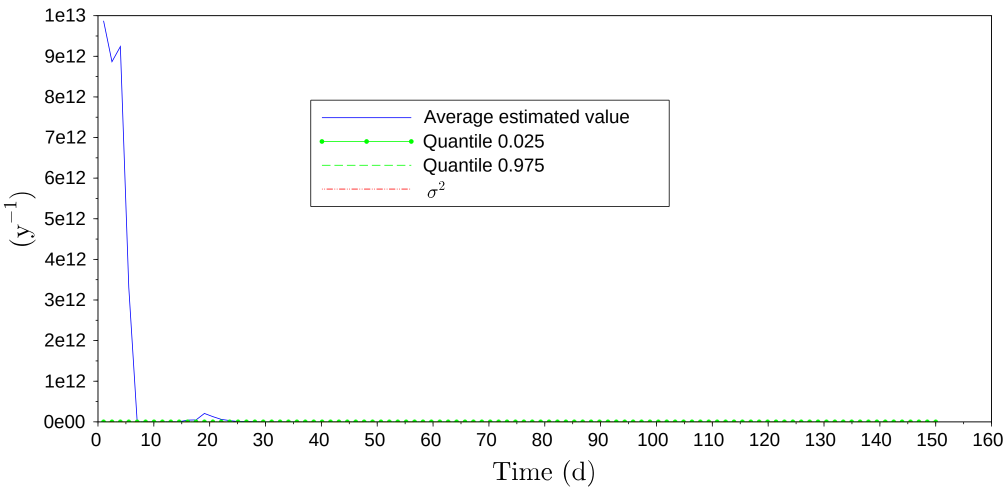

Figure 2.3 – Quantiles for the square of the short-term volatility, with 2 processes, $\vartheta = 10$ y^{-1} and deterministic constant volatilities

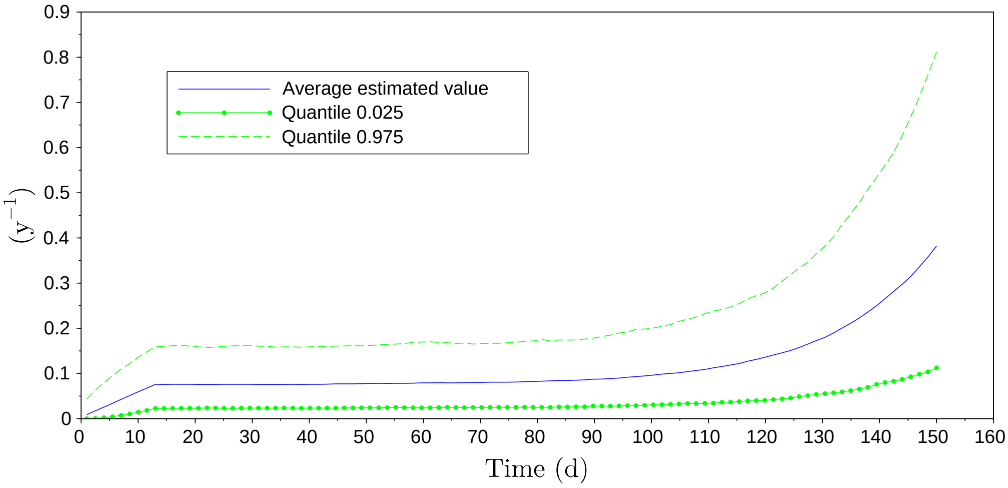

Figure 2.4 – Quantiles for the square of the equivalent volatility, with 2 processes, $\vartheta = 10$ y^{-1} and the CIR-like specification for volatility processes

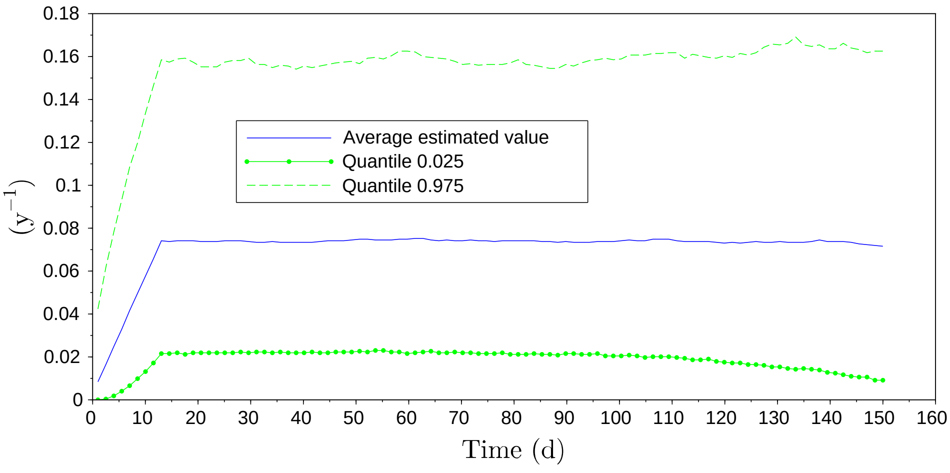

Figure 2.5 – Quantiles for the square of the long-term volatility, with 2 processes, $\vartheta = 10$ y^{-1} and the CIR-like specification for volatility processes

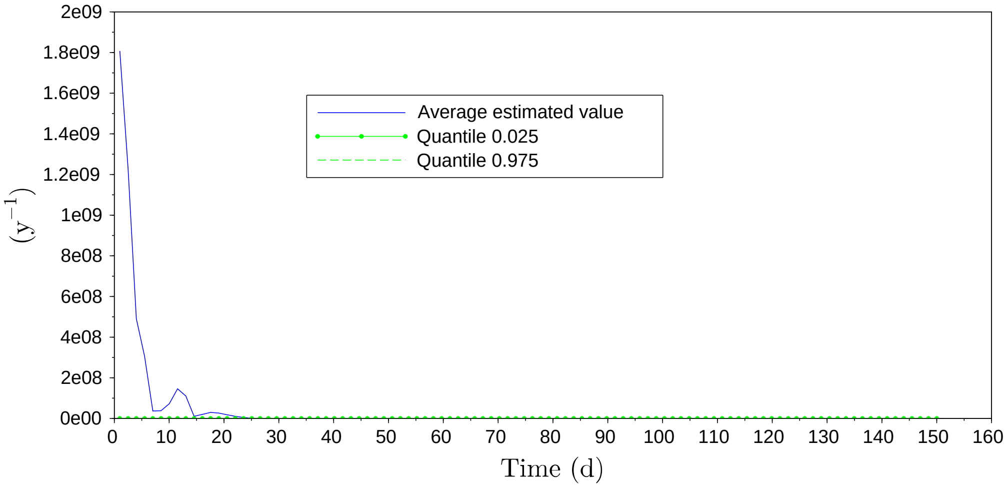

Figure 2.6 – Quantiles for the square of the short-term volatility, with 2 processes, $\vartheta = 10$ y^{-1} and the CIR-like specification for volatility processes

2.3.3 Study based on real data from the French electricity market

The data used for estimation are the 6 available month-ahead forward contracts on the French market (`www.eex.com`) from December 6th, 2001 to December 31st, 2013. On this history, we get 145 periods of 1 month ($n \simeq 20$) where 6 processes (the 6 month-ahead contracts) are jointly observed, whereas we get 141 periods of 5 months ($n \simeq 100$) where 2 processes (the 1 month-ahead and the 2 month-ahead contracts) are jointly observed. These numbers of periods are given in Table 2.6 for all the configurations described in Section 2.3.2. In the same column, Table 2.6 also precises the number of periods on which the estimator converges, convergence meaning that the value of $\hat{\vartheta}_{d,n}$ or $\tilde{\vartheta}_{2,n}$ is not zero (see the definitions of the estimators). And the same table gives the estimation results of ϑ for all the possible configurations, with the average value and the standard deviation of the estimators.

Estimator	Per. with convergence/ Number of per.	Average	Standard deviation
$\hat{\vartheta}_{2,n}$	49/141	26.065	11.788
$\tilde{\vartheta}_{2,n}$	49/141	26.081	11.779
$\hat{\vartheta}_{3,n}$	100/142	4.3707	3.5329
$\hat{\vartheta}_{4,n}$	111/143	3.1333	2.6758
$\hat{\vartheta}_{5,n}$	111/139	2.0936	2.4969
$\hat{\vartheta}_{6,n}$	105/125	3.3881	2.8221

Table 2.6 – Estimators of ϑ on real data in France (unit: y)

The main remark on these results is that, contrary to the results on simulated data, the values of the estimators are different from one configuration to another. More precisely, the estimators from 2 processes are higher (of a factor between 5 and 8) than the ones from 3 to 6 processes. This can be explained by two different causes. First, the estimators from 3 to 6 processes present a theoretical bias, of which value is unknown: this was stated in Theorem 2.1. Second, these differences may be due to the presence of errors linked to measurement or to the model.

The estimated equivalent volatility function is given in Figure 2.7, it shows an increasing volatility with respect to a decreasing time to maturity, which seems to confirm the well-known Samuelson effect.

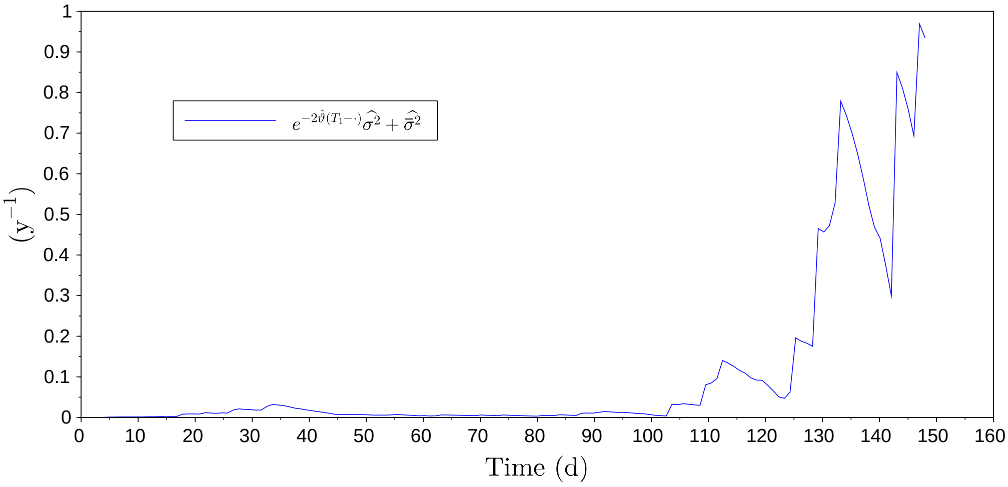

Figure 2.7 – Estimated volatilities with 2 processes observed over 5 months (February–June, 2003), $\hat{\vartheta}_{2,n} = 30.278\ \mathrm{y}^{-1}$

2.4 Proofs

2.4.1 Preliminaries: localization

Localization is a very important tool, which will in particular allow us to work under the strong assumption that the processes b, σ and $\overline{\sigma}$ are bounded, in the following proofs. Let us explain the basic idea: recall that they have been assumed to be càdlàg, and therefore they are locally bounded. This means the existence of a sequence of stopping times τ_p going to infinity such that for each p, our three processes are bounded up to time τ_p. Then we work on the event $\{\tau_p > T\}$, taking profit of the boundedness assumption. Then, as the probability of $\{\tau_p > T\}$ goes to 1 as $p \to \infty$, the results shall remain true under local boundedness assumption only.
We refer to Section 3.6.3 in [49] for a complete presentation. From now on, we will only mention, when needed, that the processes are assumed to be bounded "by localization".

2.4.2 Proof of Theorem 2.1

Proof of Theorem 2.1 (1)

Step 1 We first assume that $b^j = 0$ for $j = 1, 2$. For notational simplicity, we set $\mathfrak{e}_{\ell,k}(\vartheta) = e^{-\vartheta T_k} - e^{-\vartheta T_\ell}$. Let us define

$$\zeta_i^n = (\Delta_i^n X^2)^2 - (\Delta_i^n X^1)^2$$

and

$$\xi_i^n = \left(\Delta_i^n X^2 - \Delta_i^n X^1\right)^2.$$

Clearly

$$\begin{aligned}&\Big(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-\vartheta(T_2-t)}\sigma_t dB_t + \int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t d\overline{B}_t\Big)^2 - \Big(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-\vartheta(T_1-t)}\sigma_t dB_t + \int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t d\overline{B}_t\Big)^2 \\ =&(e^{-2\vartheta T_2} - e^{-2\vartheta T_1})\Big(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{\vartheta t}\sigma_t dB_t\Big)^2 + 2\mathfrak{e}_{1,2}(\vartheta)\int_{(i-1)\Delta_n}^{i\Delta_n} e^{\vartheta t}\sigma_t dB_t \int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t d\overline{B}_t,\end{aligned}$$

therefore, setting $\chi_i^n = 2\mathfrak{e}_{1,2}(\vartheta)\int_{(i-1)\Delta_n}^{i\Delta_n} e^{\vartheta t}\sigma_t dB_t \int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t d\overline{B}_t$, we obtain the following representation

$$\zeta_i^n = \frac{1}{\psi_{T_1,T_2}(\vartheta)}\xi_i^n + \chi_i^n. \tag{2.6}$$

By standard convergence of the quadratic variation (see for instance Section 2.1.5 in [62]),

$$\sum_{i=1}^n \xi_i^n \to \mathfrak{e}_{1,2}(\vartheta)^2 \int_0^T e^{2\vartheta t}\sigma_t^2 dt$$

in probability. Note that the limit is almost surely positive by Assumption 2.1. Also, since B and $\overline{B}$ are independent, and since $\sigma_t^2 \leq M$ and $\overline{\sigma}_t^2 \leq M$ for some constant $M > 0$ by localization, we have that

$$\mathbb{E}\Big[\Big(\sum_{i=1}^{n}\int_{(i-1)\Delta_n}^{i\Delta_n} e^{\vartheta t}\sigma t dB_t \int_{(i-1)\Delta_n}^{i\Delta_n}\overline{\sigma}_t d\overline{B}_t\Big)^2\Big]$$
$$=\sum_{i=1}^{n}\mathbb{E}\big[\big(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{\vartheta t}\sigma_t dB_t\big)^2\big]\mathbb{E}\big[\big(\int_{(i-1)\Delta_n}^{i\Delta_n}\overline{\sigma}_t d\overline{B}_t\big)^2\big] \leq \Delta_n e^{2\vartheta T}M^2 \to 0.$$

Therefore $\sum_{i=1}^{n}\zeta_i^n$ converges in probability as well, with the same limit as $\frac{1}{\psi_{T_1,T_2}(\vartheta)}\sum_{i=1}^{n}\xi_i^n$. It follows that

$$\Psi^n_{T_1,T_2} = \frac{\sum_{i=1}^{n}\big(\Delta_i^n X^2 - \Delta_i^n X^1\big)^2}{\sum_{i=1}^{n}(\Delta_i^n X^2)^2 - (\Delta_i^n X^1)^2} = \frac{\sum_{i=1}^{n}\xi_i^n}{\sum_{i=1}^{n}\zeta_i^n} \to \psi_{T_1,T_2}(\vartheta)$$

in probability. We derive the convergence

$$\psi_{T_1,T_2}(\hat{\vartheta}_{2,n}) \to \psi_{T_1,T_2}(\vartheta)$$

in probability on the event $\{\Psi^n_{T_1,T_2}\in(-1,0)\}$, hence the convergence $\hat{\vartheta}_{2,n}\to\vartheta$ in probability as well since $\{\Psi^n_{T_1,T_2}\in(-1,0)\}$ has asymptotically probability 1 and that $\vartheta \rightsquigarrow \psi_{T_1,T_2}(\vartheta)$ is invertible with continuous inverse.

Step 2 Using (2.6), we readily obtain

$$\Delta_n^{-1/2}\big(\Psi^n_{T_1,T_2}-\psi_{T_1,T_2}(\vartheta)\big) = \Delta_n^{-1/2}\Big(\frac{\sum_{i=1}^{n}\xi_i^n}{\sum_{i=1}^{n}\zeta_i^n} - \psi_{T_1,T_2}(\vartheta)\Big) = -\psi_{T_1,T_2}(\vartheta)\frac{\Delta_n^{-1/2}\sum_{i=1}^{n}\chi_i^n}{\sum_{i=1}^{n}\zeta_i^n}.$$

Consider next the sequence of 1-dimensional processes

$$\chi_n(t) = \Delta_n^{1/2}\sum_{i=1}^{\lfloor t\Delta_n^{-1}\rfloor} f\big(\Delta_n^{-1/2}\Delta_i^n Y^1, \Delta_n^{-1/2}\Delta_i^n Y^2\big),$$

where $Y_t = (Y_t^1, Y_t^2) = \big(\int_0^t e^{\vartheta s}\sigma_s dB_s, \int_0^t \overline{\sigma}_s d\overline{B}_s\big)$. By Theorem 3.21, p. 231 in [49] applied to the martingale Y with $f(x,y)=xy$ which has vanishing integral under the standard 2-dimensional-Gaussian measure, we have that the process $\chi_n(t)$ converges stably in law to a continuous process $\chi(t)$ defined on an extension of the original probability space and given by

$$\chi(t) = \int_0^t e^{\vartheta s}\sigma_s\overline{\sigma}_s dW_s,$$

where W is a Brownian motion independent of $\mathcal{F}$. Using successively $\Delta_n^{-1/2}\sum_{i=1}^{n}\chi_i^n = 2\mathfrak{c}_{1,2}(\vartheta)\chi_n(T)$, the fact that the convergence $\chi_n\to\chi$ holds stably in law and the convergence

$$\sum_{i=1}^{n}\zeta_i^n \to (e^{-2\vartheta T_2} - e^{-2\vartheta T_1})\int_0^T e^{2\vartheta t}\sigma_t^2 dt,$$

in probability, we derive

$$-\Delta_n^{-1/2}\psi_{T_1,T_2}(\vartheta)\frac{\sum_{i=1}^n \chi_i^n}{\sum_{i=1}^n \zeta_i^n} \to -\,\psi_{T_1,T_2}(\vartheta)\frac{2(e^{-\vartheta T_2}-e^{-\vartheta T_1})\chi(T)}{(e^{-2\vartheta T_2}-e^{-2\vartheta T_1})\int_0^T e^{2\vartheta t}\sigma_t^2 dt}$$
$$= -\,\frac{2(e^{-\vartheta T_2}-e^{-\vartheta T_1})^3}{(e^{-2\vartheta T_2}-e^{-2\vartheta T_1})^2\int_0^T e^{2\vartheta t}\sigma_t^2 dt}\chi(T)$$

in distribution. Conditionally to $\mathcal{F}$, the limiting variable is centered Gaussian, with conditional variance $v_\vartheta(\sigma,\overline{\sigma}) = 4\frac{(e^{-\vartheta T_2}-e^{-\vartheta T_1})^2}{(e^{-\vartheta T_2}+e^{-\vartheta T_1})^4}\frac{\int_0^T e^{2\vartheta t}\sigma_t^2\overline{\sigma}_t^2 dt}{(\int_0^T e^{2\vartheta t}\sigma_t^2 dt)^2}$.

Step 3 On the event $\{\Psi^n_{T_1,T_2} \in (-1,0)\}$, we have

$$\Delta_n^{-1/2}\big(\hat{\vartheta}_{2,n}-\vartheta\big) = \Delta_n^{-1/2}\big(\Psi^n_{T_1,T_2}-\psi_{T_1,T_2}(\vartheta)\big)\partial_\vartheta\psi^{-1}_{T_1,T_2}(Z_n)$$

for some Z_n that converges to $\psi_{T_1,T_2}(\vartheta)$ in probability by Step 1. The conclusion follows from

$$\big(\partial_\vartheta\psi^{-1}_{T_1,T_2}(\psi_{T_1,T_2}(\vartheta))\big)^2 v_\vartheta(\sigma,\overline{\sigma}) = V_\vartheta(\sigma,\overline{\sigma})$$

together with the fact that $\{\Psi^n_{T_1,T_2} \in (-1,0)\}$ has asymptotically probability 1.

Step 4 It remains to relax the restriction $b^j = 0$. When b^j is non-zero, by localization again, we may assume it is bounded. Then, by Girsanov theorem, we apply a change of measure which is $\mathcal{F}$-measurable. Since the convergence in distribution in Step 2 holds stably in law, we may work under this change of measure (see Section 2.4.4 in [62] for a simple explanation)). Finally, relaxing the boundedness assumption on $\sigma,\overline{\sigma}$ and b^j is standard, see Section 2.4.1 above.

Proof of Theorem 2.1 (2)

Step 1 We have

$$\Psi^n_{T_1,T_2,T_3} = \frac{\sum_{i=1}^n \big(\Delta_i^n(X^3-X^2)\big)^2}{\sum_{i=1}^n \big(\Delta_i^n(X^2-X^1)\big)^2}$$

By standard convergence of the quadratic variation

$$\sum_{i=1}^n \big(\Delta_i^n(X^2-X^1)\big)^2 \to \mathfrak{c}_{1,2}(\vartheta)^2\int_0^T e^{2\vartheta t}\sigma_t^2 dt, \tag{2.7}$$
$$\sum_{i=1}^n \big(\Delta_i^n(X^3-X^2)\big)^2 \to \mathfrak{c}_{2,3}(\vartheta)^2\int_0^T e^{2\vartheta t}\sigma_t^2 dt$$

in probability. Since $\psi_{T_1,T_2,T_3}(\vartheta) = \frac{\mathfrak{c}_{2,3}(\vartheta)^2}{\mathfrak{c}_{1,2}(\vartheta)^2}$, we derive $\psi_{T_1,T_2,T_3}\big(\hat{\vartheta}_{n,3}\big) \to \psi_{T_1,T_2,T_3}(\vartheta)$ in probability on the event $\big\{\Psi_{T_1,T_2,T_3} \in \big(0,\big(\frac{T_3-T_2}{T_2-T_1}\big)^2\big)\big\}$ which has asymptotically probability 1, hence the convergence $\hat{\vartheta}_{n,3} \to \vartheta$ in probability.

Step 2 We further have

$$\Psi^n_{T_1,T_2,T_3} - \psi_{T_1,T_2,T_3}(\vartheta) = \frac{\sum_{i=1}^n \big(\Delta_i^n(X^3 - X^2)\big)^2}{\sum_{i=1}^n \big(\Delta_i^n(X^2 - X^1)\big)^2} - \frac{\mathfrak{e}_{2,3}(\vartheta)^2}{\mathfrak{e}_{1,2}(\vartheta)^2} = \frac{\sum_{i=1}^n \eta_i^n}{\sum_{i=1}^n \big(\Delta_i^n(X^2 - X^1)\big)^2},$$

with

$$\eta_i^n = \big(\Delta_i^n(X^3 - X^2)\big)^2 - \frac{\mathfrak{e}_{2,3}(\vartheta)^2}{\mathfrak{e}_{1,2}(\vartheta)^2}\big(\Delta_i^n(X^2 - X^1)\big)^2.$$

Write $\overline{\Delta}_i^n f = \int_{(i-1)\Delta_n}^{i\Delta_n} f(t)dt$. One readily checks that the following decomposition holds: $\eta_i^n = (\eta')_i^n + (\eta'')_i^n$, with

$$(\eta')_i^n = \big(\overline{\Delta}_i^n(b^3 - b^2)\big)^2 - \frac{\mathfrak{e}_{2,3}(\vartheta)^2}{\mathfrak{e}_{1,2}(\vartheta)^2}\big(\overline{\Delta}_i^n(b^2 - b^1)\big)^2$$

and

$$(\eta'')_i^n = 2\mathfrak{e}_{2,3}(\vartheta)\Big(\overline{\Delta_i^n}\big((b^3 - b^2) - \frac{\mathfrak{e}_{2,3}(\vartheta)}{\mathfrak{e}_{1,2}(\vartheta)}(b^2 - b^1)\big)\Big)\int_{(i-1)\Delta_n}^{i\Delta_n} e^{\vartheta t}\sigma_t dB_t.$$

We will need the following lemma, proof of which is relatively straightforward yet technical and given in Section 2.5.1.

Lemma 2.4.1. *Let $(Y_t)_{t\geq 0}$ and $(Z_t)_{t\geq 0}$ be two càdlàg adapted processes. Assume that for some $s > 1/2$, we have $\sup_{t\in[0,T]} t^{-s}\omega(Y)_t < \infty$. Then*

$$\Delta_n^{-1}\sum_{i=1}^n \big(\overline{\Delta}_i^n Y\big)^2 \to \int_0^T Y_t^2 dt$$

and

$$\Delta_n^{-1}\sum_{i=1}^n \overline{\Delta}_i^n(Y)\int_{(i-1)\Delta_n}^{i\Delta_n} Z_t dB_t \to \int_0^T Y_t Z_t dB_t$$

in probability.

We successively have

$$\Delta_n^{-1}\sum_{i=1}^n (\eta')_i^n \to \int_0^T \mu_\vartheta(b_t)dt$$

with $\mu_\vartheta(b_t) = (b_t^3 - b_t^2)^2 - \frac{\mathfrak{e}_{2,3}(\vartheta)^2}{\mathfrak{e}_{1,2}(\vartheta)^2}(b_t^2 - b_t^1)^2$ and

$$\Delta_n^{-1}\sum_{i=1}^n (\eta'')_i^n \to 2\int_0^T \lambda_\vartheta(b_t)e^{\vartheta t}\sigma_t dB_t$$

in probability, by Lemma 2.4.1 applied to $Y_t = (b_t^3 - b_t^2) - \frac{\mathfrak{e}_{2,3}(\vartheta)}{\mathfrak{e}_{1,2}(\vartheta)}(b_t^2 - b_t^1)$ and $Z_t = e^{\vartheta t}\sigma_t$, and Assumption 2.1, where $\lambda_\vartheta(b_t) = \mathfrak{e}_{2,3}(\vartheta)Y_t$. This, together with (2.7), implies the convergence

$$\Delta_n^{-1}\big(\Psi^n_{T_1,T_2,T_3} - \psi_{T_1,T_2,T_3}(\vartheta)\big) \to \frac{\int_0^T \mu_\vartheta(b_t)dt + 2\int_0^T \lambda_\vartheta(b_t)e^{\vartheta t}\sigma_t dB_t}{\mathfrak{e}_{1,2}(\vartheta)^2\int_0^T e^{2\vartheta t}\sigma_t^2 dt}$$

in probability.

Step 3 Finally, we have

$$\Delta_n^{-1}\big(\hat{\vartheta}_{3,n}-\vartheta\big)=\Delta_n^{-1}\big(\Psi^n_{T_1,T_2,T_3}-\psi_{T_1,T_2,T_3}(\vartheta)\big)\partial_\vartheta\psi^{-1}_{T_1,T_2,T_3}(Z_n),$$

for some Z_n that converges to $\psi_{T_1,T_2,T_3}(\vartheta)$ by Step 1. Hence

$$\Delta_n^{-1}\big(\hat{\vartheta}_{3,n}-\vartheta\big)\to\frac{\int_0^T\mu_\vartheta(b_t)dt+2\int_0^T\lambda_\vartheta(b_t)e^{\vartheta t}\sigma_t dB_t}{\partial_\vartheta\psi_{T_1,T_2,T_3}(\vartheta)\mathfrak{c}_{1,2}(\vartheta)^2\int_0^T e^{2\vartheta t}\sigma_t^2dt}$$

and we conclude by noting that $\partial_\vartheta\psi_{T_1,T_2,T_3}(\vartheta)=\frac{2D_3}{\mathfrak{c}_{1,2}(\vartheta)^3}$.

2.4.3 Proof of Theorem 2.2

We shall examine the rates of convergence of the two estimators $\widehat{\sigma}^2_{n,t}$ and $\widehat{\widehat{\sigma}}^2_{n,t}$ separately. The proof is led while assuming that $b^1=b^2=0$. For ease of notation, we write $\hat{\vartheta}_{2,n}$ for $\max\{\hat{\vartheta}_{2,n},\varpi_n\}$ and set $t_i=i\Delta_n$ for $i=1,\dots,n$. We also define $K(t)=\mathbf{1}_{(0,1]}(t)$ and $K_h(t)=h^{-1}K(th^{-1})$ for $h>0$. We have

$$\widehat{\sigma}^2_{n,t}-\sigma_t^2=\frac{\sum_{i=1}^n K_{h_n}\big(t-t_{i-1}\big)\big((\Delta_i^nX^1)^2-(\Delta_i^nX^2)^2\big)}{e^{-2\hat{\vartheta}_{2,n}(T_1-t)}-e^{-2\hat{\vartheta}_{2,n}(T_2-t)}}-\sigma_t^2=I+II,$$

with

$$\begin{aligned}I=&\Big(\frac{1}{e^{-2\hat{\vartheta}_{2,n}(T_1-t)}-e^{-2\hat{\vartheta}_{2,n}(T_2-t)}}-\frac{1}{e^{-2\vartheta(T_1-t)}-e^{-2\vartheta(T_2-t)}}\Big)\\&\times\sum_{i=1}^n K_{h_n}\big(t-t_{i-1}\big)\big((\Delta_i^nX^1)^2-(\Delta_i^nX^2)^2\big)\end{aligned}$$

and

$$II=\frac{\sum_{i=1}^n K_{h_n}\big(t-t_{i-1}\big)\big((\Delta_i^nX^1)^2-(\Delta_i^nX^2)^2\big)}{e^{-2\vartheta(T_1-t)}-e^{-2\vartheta(T_2-t)}}-\sigma_t^2.$$

The term I Since $\mathbb{E}[\big(\Delta_i^nX^j\big)^2]$ is of order Δ_n by Burkholder-Davis-Gundy inequality, we have that $\mathbb{E}[\big|(\Delta_i^nX^1)^2-(\Delta_i^nX^2)^2\big|]$ is of order Δ_n as well and therefore

$$\begin{aligned}&\mathbb{E}\Big[\Big|\sum_{i=1}^n K_{h_n}(t-t_{i-1})\big((\Delta_i^nX^1)^2-(\Delta_i^nX^2)^2\big)\Big|\Big]\\\leq&\sum_{i=1}^n K_{h_n}(t-t_{i-1})\mathbb{E}\Big[\big|(\Delta_i^nX^1)^2-(\Delta_i^nX^2)^2\big|\Big]\\\lesssim&\sum_{i=1}^n K_{h_n}(t-t_{i-1})\Delta_n\lesssim 1\end{aligned}$$

since $K_{h_n}(t-t_{i-1})$ is of order h_n^{-1} for a number of terms that are at most of order $\Delta_n^{-1}h_n$. Therefore $\sum_{i=1}^n K_{h_n}\big(t-t_{i-1}\big)\big((\Delta_i^nX^1)^2-(\Delta_i^nX^2)^2\big)$ is bounded in expectation hence tight, and we conclude that I is of order $\Delta_n^{1/2}$ in probability by applying Theorem 2.1 (1).

The term II The term II further splits into $II = (e^{-2\vartheta(T_1-t)} - e^{-2\vartheta(T_2-t)})^{-1}\big(B_n(t)+V_n(t)\big)$, having

$$V_n(t) = \sum_{i=1}^{n} K_{h_n}(t-t_{i-1})\big((\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2 - \mathbb{E}\big[(\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2 \big| \mathcal{F}_{i-1}\big]\big)$$

and

$$B_n(t) = \sum_{i=1}^{n} \mathbb{E}\big[K_{h_n}(t-t_{i-1})\big((\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2\big)\big|\mathcal{F}_{i-1}\big] - \big(e^{-2\vartheta(T_1-t)} - e^{-2\vartheta(T_2-t)}\big)\sigma_t^2.$$

Here and in what follows, we use the shorter notation $\mathcal{F}_i$ for $\mathcal{F}_{t_i}$.

Bounding the variance term

We first prove an upper bound for $\mathbb{E}(V_n(t)^2)$ uniformly in $t \in [h_n, T]$. We have

$$\begin{aligned}
&\sup_{t\in[h_n,T]} \mathbb{E}\Big[\Big(\sum_{i=1}^{n} K_{h_n}(t-t_{i-1})\big((\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2 - \mathbb{E}\big[(\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2 \big| \mathcal{F}_{i-1}\big]\big)\Big)^2\Big] \\
&= \sup_{t\in[h_n,T]} \mathbb{E}\Big[\Big(\sum_{i=1}^{n} K_{h_n}(t-t_{i-1})^2\Big((\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2 - \mathbb{E}\big[(\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2 \big| \mathcal{F}_{i-1}\big]\Big)^2\Big)\Big] \\
&= \sup_{t\in[h_n,T]} h_n^{-2}\sum_{i=1}^{n} K^2\Big(\frac{t-t_{i-1}}{h_n}\Big)\mathbb{E}\Big[\big((\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2 - \mathbb{E}\big[(\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2 \big| \mathcal{F}_{i-1}\big]\big)^2\Big]
\end{aligned}$$

because cross-terms in the development of the second-line are zero due to conditioning. Then, there are at most $O(\Delta_n^{-1} h_n)$ terms that are not zero in the sum, because K has compact support, and this estimate is uniform in $t \in [h_n, T]$. Now, since K is bounded, and because

$$\mathbb{E}\Big[\big((\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2 - \mathbb{E}\big((\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2 \big| \mathcal{F}_{i-1}\big)\big)^2\Big] \lesssim \Delta_n^2,$$

we obtain $\sup_{t\in[h_n,T]} \mathbb{E}\big[(V_n(t))^2\big] \lesssim \Delta_n h_n^{-1}$.

Bounding the bias term

In order to get an upper bound for the bias term $\sup_{t\in[h_n,T]} \mathbb{E}\big((B_n(t))^2\big)$, we use the decomposition

$$B_n(t) = \big(e^{-2\vartheta(T_1-t)} - e^{-2\vartheta(T_2-t)}\big)(B_I(t) + B_{II}(t)),$$

where

$$B_I(t) = \int_0^T h_n^{-1} K\Big(\frac{t-u}{h_n}\Big) e^{-2\vartheta(t-u)}\sigma_u^2 du - \sigma_t^2$$

and

$$B_{II}(t) = \frac{\sum_{i=1}^{n}\mathbb{E}\Big(h_n^{-1}K\Big(\frac{t-t_{i-1}}{h_n}\Big)\big((\Delta_i^n X^1)^2-(\Delta_i^n X^2)^2\big)\Big|\mathcal{F}_{i-1}\Big)}{e^{-2\vartheta(T_1-t)}-e^{-2\vartheta(T_2-t)}} - \int_0^T h_n^{-1}K\Big(\frac{t-u}{h_n}\Big)e^{-2\vartheta(t-u)}\sigma_u^2 du.$$

First we treat the term B_I. We have for every $t\in[h_n,T]$, the property $\int_{\frac{t-T}{h_n}}^{\frac{t}{h_n}} K(x)dx=1$ and thus

$$\sup_{t\in[h_n,T]}\mathbb{E}\big((B_I(t))^2\big) = \sup_{t\in[h_n,T]}\mathbb{E}\Big[\Big(\int_0^T h_n^{-1}K\Big(\frac{t-u}{h_n}\Big)e^{-2\vartheta(t-u)}\sigma_u^2du-\sigma_t^2\Big)^2\Big]$$
$$= \sup_{t\in[h_n,T]}\mathbb{E}\Big[\Big(\int_{\frac{t-T}{h_n}}^{\frac{t}{h_n}} K(x)e^{-2\vartheta h_n x}\sigma_{t-h_nx}^2dx-\sigma_t^2\Big)^2\Big].$$

We have $\mathrm{supp}(K)\subset\big[\frac{t-T}{h},\frac{t}{h}\big]$. Furthermore, the Lebesgue measure of the support of K is 1, so that

$$\sup_{t\in[h_n,T]}\mathbb{E}\big((B_I(t))^2\big) = \sup_{t\in[h_n,T]}\mathbb{E}\Big[\Big(\int_{\mathrm{supp}(K)} K(x)\big(e^{-2\vartheta h_nx}\sigma_{t-h_nx}^2-\sigma_t^2\big)dx\Big)^2\Big]$$
$$\le \sup_{t\in[h_n,T]}\int_{\mathrm{supp}(K)} K^2(x)\mathbb{E}\big[\big(e^{-2\vartheta h_nx}\sigma_{t-h_nx}^2-\sigma_t^2\big)^2\big]dx$$

using Jensen inequality. Then, by convexity inequality,

$$(e^{-2\vartheta h_nx}\sigma_{t-h_nx}^2-\sigma_t^2)^2\le 2(e^{-2\vartheta h_nx}\sigma_{t-h_nx}^2-\sigma_{t-h_nx}^2)^2+2(\sigma_{t-h_nx}^2-\sigma_t^2)^2.$$

By bounding the Lagrange remainder in the Taylor series at order 0 of $x\mapsto e^{-2\vartheta h_nx}$ at the point 0, we get $|e^{-2\vartheta h_nx}-1|\le M|2\vartheta h_nx|$ for some $M>0$. By localization, we find some $M_\sigma>0$ such that $\sigma_t<M_\sigma$. There remains

$$\sup_{t\in[h_n,T]}\mathbb{E}\big((B_I(t))^2\big)\le \sup_{t\in[h_n,T]}\int_{\mathrm{supp}(K)}K^2(x)\big(2M_\sigma^4(2\vartheta h_nxM)^2+2\mathbb{E}\big((\sigma_{t-h_nx}^2-\sigma_t^2)^2\big)\big)dx$$
$$\le \int_{\mathrm{supp}(K)}K^2(x)\big(2M_\sigma^4(2\vartheta h_nxM)^2+2c|h_nx|^{2\alpha}\big)dx,$$

using Assumption 2.2. Therefore

$$\sup_{t\in[h_n,T]}\mathbb{E}\big((B_I(t))^2\big)\lesssim h_n^{2\alpha}.$$

Let us now bound the bias term B_{II}. We have

$$B_{II}(t)=\bar{B}_{II}(t)+\tilde{B}_{II}(t),$$

where

$$\bar{B}_{II}(t) = \sum_{i=1}^{n} h_n^{-1}\delta_i(t)$$

with

$$\delta_i(t) = \mathbb{E}\Big(K\Big(\frac{t-t_{i-1}}{h_n}\Big)\int_{t_{i-1}}^{t_i} e^{-2\vartheta(t-u)}\sigma_u^2 du\Big|\mathcal{F}_{i-1}\Big) - \int_{t_{i-1}}^{t_i} K\Big(\frac{t-u}{h_n}\Big)e^{-2\vartheta(t-u)}\sigma_u^2 du,$$

and

$$\begin{aligned}
\tilde{B}_{II}(t) &= \sum_{i=1}^{n} h_n^{-1}K\Big(\frac{t-t_{i-1}}{h_n}\Big)\mathbb{E}\Big(\frac{\big((\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2\big)}{e^{-2\vartheta(T_1-t)} - e^{-2\vartheta(T_2-t)}} - \int_{t_{i-1}}^{t_i} e^{-2\vartheta(t-u)}\sigma_u^2 du\Big|\mathcal{F}_{i-1}\Big)\\
&= \sum_{i=1}^{n} h_n^{-1}K\Big(\frac{t-t_{i-1}}{h_n}\Big)e^{-2\vartheta t}\mathbb{E}\Big(\Big[\int_{t_{i-1}}^{t_i} e^{\vartheta u}\sigma_u dB_u\Big]^2 - \int_{t_{i-1}}^{t_i} e^{2\vartheta u}\sigma_u^2 du\Big|\mathcal{F}_{i-1}\Big)\\
&= 0.
\end{aligned}$$

Also,

$$\begin{aligned}
\sup_{t\in[h_n,T]} \mathbb{E}\big((B_{II}(t))^2\big) &= \sup_{t\in[h_n,T]} \mathbb{E}\big((\bar{B}_{II}(t))^2\big)\\
&= \sup_{t\in[h_n,T]} \sum_{i=1}^{n} \mathbb{E}\Big(h_n^{-2}(\delta_i(t))^2\Big) + 2\sup_{t\in[h_n,T]} \sum_{1\le i<j\le n} \mathbb{E}\Big(h_n^{-2}\delta_i(t)\delta_j(t)\Big).
\end{aligned}$$

We have

$$\mathbb{E}\Big(h_n^{-2}(\delta_i^{\mathcal{S}}(t))^2\Big) \le \frac{\Delta_n}{h_n^2}\int_{t_{i-1}}^{t_i} e^{-4\vartheta(t-u)}\mathbb{E}\Big(\Big(K\Big(\frac{t-t_{i-1}}{h_n}\Big)\mathbb{E}(\sigma_u^2|\mathcal{F}_{i-1}) - K\Big(\frac{t-u}{h_n}\Big)\sigma_u^2\Big)^2\Big)$$

by Jensen inequality, so that $\sum_{i=1}^n \mathbb{E}\Big(h_n^{-2}(\delta_i^{\mathcal{S}}(t))^2\Big) \lesssim \Delta_n h_n^{-1}$ uniformly on t, as K is bounded and there are at most $O(\Delta_n h_n)$ terms that are not zero in the sum. Then, by conditioning on $\mathcal{F}_{j-1}$,

$$\mathbb{E}\Big(h_n^{-2}\delta_i(t)\delta_j(t)\Big) = h_n^{-2}\mathbb{E}\Big(\delta_i(t)\mathbb{E}\Big(\int_{t_{j-1}}^{t_j}\Big(K\Big(\frac{t-t_{j-1}}{h_n}\Big) - K\Big(\frac{t-u}{h_n}\Big)\Big)e^{-2\vartheta(t-u)}\sigma_u^2 du\Big|\mathcal{F}_{j-1}\Big)\Big),$$

and the difference $K\Big(\frac{t-t_{j-1}}{h_n}\Big) - K\Big(\frac{t-u}{h_n}\Big)$ is non-zero only if $t\in(t_{j-1},u]$ or $t\in(t_{j-1}+h_n, u+h_n]$, which can be the case for j in some set $\mathcal{J}_t$, which contains at most three indexes. Therefore,

$$\begin{aligned}
\Big|\sum_{1\le i<j\le n} \mathbb{E}\Big(h_n^{-2}\delta_i(t)\delta_j(t)\Big)\Big| &= \Big|\sum_{i=1}^{n-1}\sum_{j\in\mathcal{J}_t} \mathbb{E}\Big(h_n^{-2}\delta_i(t)\delta_j(t)\Big)\Big|\\
&\le 3h_n^{-2}\sum_{i=1}^{n-1}\mathbb{E}\big(|\delta_i(t)|\big)M_\sigma^2 e^{2\vartheta T}\Delta_n,
\end{aligned}$$

which is of order $\Delta_n h_n^{-1}$, for the same reasons than before. We can thus conclude that

$$\sup_{t\in[h_n,T]} \mathbb{E}\big((B_{II}(t))^2\big) \lesssim \Delta_n h_n^{-1}.$$

As we got previously $\sup_{t\in[h_n,T]} \mathbb{E}\big[(V_n(t))^2\big] \lesssim \Delta_n h_n^{-1}$, $\sup_{t\in[h_n,T]} \mathbb{E}\big((B_I(t))^2\big) \lesssim h_n^{2\alpha}$ and $\sup_{t\in[h_n,T]} \mathbb{E}\big((B_{II}(t))^2\big) \lesssim \Delta_n h_n^{-1}$, the choice $h_n = \Delta_n^{1/(2\alpha+1)}$ implies that the two error terms $h_n^{2\alpha}$ and $\Delta_n h_n^{-1}$ are of the same order, namely $\Delta_n^{2\alpha/(2\alpha+1)}$, which ends the proof concerning the volatility process σ.

To get the same result for $\widehat{\overline{\sigma}}^2$, we split $\widehat{\overline{\sigma}}^2_{n,t} - \overline{\sigma}_t^2$ in this way:

$$\begin{aligned}
\widehat{\overline{\sigma}}^2_{n,t} - \overline{\sigma}_t^2 =& \Big(\frac{e^{-2\hat{\vartheta}_{2,n}T_2}}{e^{-2\hat{\vartheta}_{2,n}T_1} - e^{-2\hat{\vartheta}_{2,n}T_2}} - \frac{e^{-2\vartheta T_2}}{e^{-2\vartheta T_1} - e^{-2\vartheta T_2}}\Big) \sum_{i=1}^n h_n^{-1} K\Big(\frac{t-t_{i-1}}{h_n}\Big)(\Delta_i^n X^1)^2 \\
&+ \Big(\frac{e^{-2\hat{\vartheta}_{2,n}T_1}}{e^{-2\hat{\vartheta}_{2,n}T_1} - e^{-2\hat{\vartheta}_{2,n}T_2}} - \frac{e^{-2\vartheta T_1}}{e^{-2\vartheta T_1} - e^{-2\vartheta T_2}}\Big) \sum_{i=1}^n h_n^{-1} K\Big(\frac{t-t_{i-1}}{h_n}\Big)(\Delta_i^n X^2)^2 \\
&+ \frac{\sum_{i=1}^n h_n^{-1} K\Big(\frac{t-t_{i-1}}{h_n}\Big)(e^{-2\vartheta(T_1-t)}(\Delta_i^n X^2)^2 - e^{-2\vartheta(T_2-t)}(\Delta_i^n X^1)^2)}{e^{-2\vartheta(T_1-t)} - e^{-2\vartheta(T_2-t)}} - \overline{\sigma}_t^2.
\end{aligned}$$

Then we proceed in the same way to find that with the choice $h_n = \Delta_n^{1/(2\alpha+1)}$, the error terms are of order $\Delta_n^{2\alpha/(2\alpha+1)}$.

This proves that the sequence

$$\Delta_n^{\alpha/(2\alpha+1)} \sup_{t\in[h_n,T]} \Big[\big|\widehat{\sigma}^2_{n,t} - \sigma_t^2\big| + \big|\widehat{\overline{\sigma}}^2_{n,t} - \overline{\sigma}_t^2\big|\Big]$$

is tight, which is also true for $t \in \mathcal{D}$, $\mathcal{D}$ being any compact interval included in $(0,T]$, as it will be included in all intervals $[h_n, T]$ for n high enough.

Adding a non-zero drift does not change the result, by the usual argument based on Girsanov theorem. Adopting local boundedness only is also done in the standard way.

2.4.4 Proof of Theorem 2.3

In this proof, we do as if $b^1 = b^2 = 0$. This simplyfing assumption may be removed afterwards, as we already did it.

We have to introduce the framework necessary for the calculus of a lower bound for the variance. All the material can be found in Sections 25.3–25.4 of [73].
Assume we observe $(\Delta_i^n X^1, \Delta_i^n X^2)$, $i = 1, \ldots, n$, having

$$\begin{aligned}
\begin{pmatrix}\Delta_i^n X^1\\ \Delta_i^n X^2\end{pmatrix} &= \begin{pmatrix}\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-\vartheta(T_1-t)}\sigma_t dB_t + \int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t d\overline{B}_t \\ \int_{(i-1)\Delta_n}^{i\Delta_n} e^{-\vartheta(T_2-t)}\sigma_t dB_t + \int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t d\overline{B}_t\end{pmatrix} \\
&= \begin{pmatrix}\big(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)}\sigma_t^2 dt\big)^{1/2}\epsilon_i + \big(\int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t^2 dt\big)^{1/2}\tilde{\epsilon}_i \\ e^{-\vartheta(T_2-T_1)}\big(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)}\sigma_t^2 dt\big)^{1/2}\epsilon_i + \big(\int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t^2 dt\big)^{1/2}\tilde{\epsilon}_i\end{pmatrix}
\end{aligned}$$

where $\epsilon_i \sim \mathcal{N}(0,1)$ and $\tilde{\epsilon}_i \sim \mathcal{N}(0,1)$ are independent, because we have Wiener integrals due to the fact that the volatility processes are deterministic. The density wrt Lebesgue measure $dxdy$ on $\mathbb{R}^2$ of the i^{th} observation is $f^i_{\vartheta,\sigma,\overline{\sigma}}$, given by

$$(x,y) \mapsto \frac{\exp\Big(- \frac{\big(\int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t^2 dt\big)(x-y)^2 + \big(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)}\sigma_t^2 dt\big)\big(y - e^{-\vartheta(T_2-T_1)}x\big)^2}{2\big(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)}\sigma_t^2 dt\big)\big(\int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t^2 dt\big)\big(1-e^{-\vartheta(T_2-T_1)}\big)^2}\Big)}{2\pi\big(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)}\sigma_t^2 dt\big)^{1/2}\big(\int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t^2 dt\big)^{1/2}\big(1-e^{-\vartheta(T_2-T_1)}\big)}. \tag{2.8}$$

The notation $f^i_{\vartheta,\sigma,\overline{\sigma}}$ underlines the fact that the density is related to the parameter of interest ϑ, which lies in $(0,+\infty)$, and to $(\sigma,\overline{\sigma})$, which is a nuisance parameter in an infinite-dimensional space. If $(\sigma,\overline{\sigma})$ were known, we would perform maximum likelihood estimation by solving the score equation:

$$\frac{1}{n}\sum_{i=1}^{n} \ell^i_{\vartheta,\sigma,\overline{\sigma}} = 0$$

with $\ell^i_{\vartheta,\sigma,\overline{\sigma}} = \partial_\vartheta \log f^i_{\vartheta,\sigma,\overline{\sigma}}(\Delta_i^n X^1, \Delta_i^n X^2) \in L_2(\mathbb{P}_{\vartheta,\sigma,\overline{\sigma}})$, which is the set of measurable functions g with $\int g^2 d\mathbb{P}_{\vartheta,\sigma,\overline{\sigma}} < \infty$, and we would be able to perform efficient estimation, with Fisher information equivalent to $\frac{1}{n}\sum_{i=1}^n \mathbb{E}((\ell^i_{\vartheta,\sigma,\overline{\sigma}})^2)$. Unfortunately, this somewhat idealistic information will not be attained as we do not know the volatility functions σ and $\overline{\sigma}$.

From the density given by Equation (2.8), we can derive $\ell^i_{\vartheta,\sigma,\overline{\sigma}}$, of which expression is

$$\begin{aligned}
&\frac{\int_{(i-1)\Delta_n}^{i\Delta_n}(T_1-t)e^{-2\vartheta(T_1-t)}\sigma_t^2 dt}{\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)}\sigma_t^2 dt} - \frac{e^{-\vartheta(T_2-T_1)}(T_2-T_1)}{1-e^{-\vartheta(T_2-T_1)}} \\
&+(\Delta_i^n X^2 - \Delta_i^n X^1)^2 \frac{\int_{(i-1)\Delta_n}^{i\Delta_n}(e^{-\vartheta(T_2-T_1)}(T_2-t)-(T_1-t))e^{-2\vartheta(T_1-t)}\sigma_t^2 dt}{\big(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)}\sigma_t^2 dt)^2(1-e^{-\vartheta(T_2-T_1)})^3} \\
&+(\Delta_i^n X^2 - e^{-\vartheta(T_2-T_1)}\Delta_i^n X^1)^2 \frac{(T_2-T_1)e^{-\vartheta(T_2-T_1)}}{\int_{(i-1)\Delta_n}^{i\Delta_n}\overline{\sigma}_t^2 dt(1-e^{-\vartheta(T_2-T_1)})^3} \\
&-\Delta_i^n X^1(\Delta_i^n X^2 - e^{-\vartheta(T_2-T_1)}\Delta_i^n X^1)\frac{e^{-\vartheta(T_2-T_1)}(T_2-T_1)}{\int_{(i-1)\Delta_n}^{i\Delta_n}\overline{\sigma}_t^2 dt(1-e^{-\vartheta(T_2-T_1)})^2}.
\end{aligned}$$

To get the best reachable information in the model

$$\mathcal{P}_i = (f^i_{\vartheta,\sigma,\overline{\sigma}})_{\vartheta\in(0,+\infty),(\sigma,\overline{\sigma})\in\Sigma(c,\tilde{c})},$$

we have to consider a parametric submodel of the form $\mathcal{P}_i^0 = (f^i_{\vartheta+\iota u,\eta^u,\overline{\eta}^u})_{0\le u\le\varepsilon}$, for $\iota \in \mathbb{R}$ and functions η^u and $\overline{\eta}^u$. $\varepsilon > 0$ is such that $\vartheta + \iota\varepsilon > 0$ and $\forall u \in [0,\varepsilon], \forall t \in [0,T], \tilde{c} < \eta_t^u < c$ and $\tilde{c} < \overline{\eta}_t^u < c$. Furthermore, we impose that $\forall t \in [0,T], \eta_t^0 = \sigma_t^2$ and $\overline{\eta}_t^0 = \overline{\sigma}_t^2$.
Note that such submodel $\mathcal{P}_i^0$ passes through the true distribution (for $u = 0$). We consider only submodels that are differentiable in quadratic mean at $u = 0$, with score function

$g_{i,\iota,\eta,\overline{\eta}} \in L_2(\mathbb{P}_{\vartheta,\sigma,\overline{\sigma}})$. If we let $\mathcal{P}_i^0$ range over all admissible submodels, then we will get a collection of score functions, which define the tangent set $\dot{\mathcal{P}}_{i,\vartheta,\sigma,\overline{\sigma}}$ of the model $\mathcal{P}_i$ at the true distribution.

Using an admissible map $u \mapsto (\eta^u, \overline{\eta}^u)$ on $[0,\varepsilon]$ and $\iota \in \mathbb{R}$, the score function $g_{i,\iota,\eta,\overline{\eta}}$ may be written as

$$g_{i,\iota,\eta,\overline{\eta}} = \iota \ell^i_{\vartheta,\sigma,\overline{\sigma}} + g_{i,\eta,\overline{\eta}},$$

where $\ell^i_{\vartheta,\sigma,\overline{\sigma}}$ is the score function when σ and $\overline{\sigma}$ are known, and $g_{i,\eta,\overline{\eta}}$ is the score function got from a parametric submodel with the parameter $\iota = 0$, which is to be interpreted as a score function related to the nuisance parameter only, while $\ell^i_{\vartheta,\sigma,\overline{\sigma}}$ corresponds to the parameter of interest.

Here we consider a parametric submodel $\mathcal{P}_i^0 = (f^i_{\vartheta+\iota u,\eta^u,\overline{\eta}^u})_{0\le u\le\varepsilon}$, where $\iota \in \mathbb{R}$, and for $t \in [0,T]$,

$$\eta^u_t = (1+uk(t))\sigma_t \text{ and } \overline{\eta}^u_t = (1+u\overline{k}(t))\overline{\sigma}_t.$$

The parameter $\varepsilon > 0$ is chosen such that $\vartheta + \iota\varepsilon > 0$ and $\forall u \in [0,\varepsilon], (\eta^u, \overline{\eta}^u) \in \Sigma(c,\tilde{c})$. It is enough to consider this simple submodel, as only the terms of order 0 and 1 of the Taylor expansions in u of η and $\overline{\eta}$ matter while defining the tangent set.
The submodel is differentiable in quadratic mean at $u = 0$, with score function $g_{i,\iota,\eta,\overline{\eta}} = g_{i,\iota,k,\overline{k}}$, which we write as

$$g_{i,\iota,k,\overline{k}} = \iota \ell^i_{\vartheta,\sigma,\overline{\sigma}} + g_{i,k,\overline{k}}.$$

Formally, $\iota\ell^i_{\vartheta,\sigma,\overline{\sigma}}$ was got as $\frac{d}{du}\Big|_{u=0} \log f^i_{\vartheta+\iota u,\sigma,\overline{\sigma}}$. On the same way, as $g_{i,k,\overline{k}}$ is relative to the nuisance parameter, it is given by $\frac{d}{du}\Big|_{u=0} \log f^i_{\vartheta,\eta^u,\overline{\eta}^u}$. That is,

$$\begin{aligned} g_{i,k,\overline{k}} = &- \frac{\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)}\sigma_t^2 k(t)dt}{\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)}\sigma_t^2 dt} - \frac{\int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t^2\overline{k}(t)dt}{\int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t^2 dt} \\ &+ (\Delta_i^n X^2 - \Delta_i^n X^1)^2 \frac{\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)}\sigma_t^2 k(t)dt}{\left(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)}\sigma_t^2 dt\right)^2 (1-e^{-\vartheta(T_2-T_1)})^2} \\ &+ (\Delta_i^n X^2 - e^{-\vartheta(T_2-T_1)}\Delta_i^n X^1)^2 \frac{\int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t^2\overline{k}(t)dt}{\left(\int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t^2 dt\right)^2 (1-e^{-\vartheta(T_2-T_1)})^2}. \end{aligned}$$

Let us introduce the operator Π for the orthogonal projection onto the closure in $L_2(\mathbb{P}_{\vartheta,\sigma,\overline{\sigma}})$ of the linear span of the tangent set for the nuisance parameter $(\sigma,\overline{\sigma})$, which is the set of all functions $g_{i,k,\overline{k}}$. Then $\widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}} = l^i_{\vartheta,\sigma,\overline{\sigma}} - \Pi l^i_{\vartheta,\sigma,\overline{\sigma}}$ is called the efficient score for ϑ. Furthermore, $\widetilde{I}_{i,\vartheta,\sigma,\overline{\sigma}} = \int (\widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}})^2 d\mathbb{P}_{\vartheta,\sigma,\overline{\sigma}}$ is called the efficient information.
We may write orthogonality conditions in order to compute $\Pi\ell^i_{\vartheta,\sigma,\overline{\sigma}}$. More precisely, it should satisfy

$$\langle \ell^i_{\vartheta,\sigma,\overline{\sigma}} - \Pi\ell^i_{\vartheta,\sigma,\overline{\sigma}}, g_{i,k,\overline{k}} \rangle = 0$$

for all functions k and $\overline{k}$. If we further expect that $\Pi\ell^i_{\vartheta,\sigma,\overline{\sigma}}$ will be some $g_{i,k^*,\overline{k}^*}$, then we just have to look for functions k^* and $\overline{k}^*$ such that for all functions k and $\overline{k}$,

$$0 = \langle \ell^i_{\vartheta,\sigma,\overline{\sigma}} - g_{i,k^*,\overline{k}^*}, g_{i,k,\overline{k}} \rangle = \int \left(l^i_{\vartheta,\sigma,\overline{\sigma}} - g_{i,k^*,\overline{k}^*} \right) g_{i,k,\overline{k}} d\mathbb{P}_{\vartheta,\sigma,\overline{\sigma}}.$$

After some computations, we find that $k^*(t) = \frac{(T_2-t)e^{-\vartheta(T_2-T_1)}t(T_1-t)}{1-e^{-\vartheta(T_2-T_1)}}$ and $\overline{k}^*(t) = 0$ satisfy that condition. The efficient score function $\widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}}$ is thus

$$\ell^i_{\vartheta,\sigma,\overline{\sigma}} - g_{i,k^*,\overline{k}^*} = \frac{(\Delta^n_i X^2 - \Delta^n_i X^1)(\Delta^n_i X^2 - e^{-\vartheta(T_2-T_1)}\Delta^n_i X^1)e^{-\vartheta(T_2-T_1)}(T_2 - T_1)}{(1 - e^{-\vartheta(T_2-T_1)})^3 \int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}^2_t dt}.$$

The expression of the efficient information follows:

$$\widetilde{I}_{i,\vartheta,\sigma,\overline{\sigma}} = \frac{(T_2 - T_1)^2}{(e^{\vartheta(T_2-T_1)} - 1)^2} \frac{\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)}\sigma^2_t dt}{\int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}^2_t dt}.$$

Using the independence of the couples of observations, we can sum $\widetilde{I}_{i,\vartheta,\sigma,\overline{\sigma}}$ over $i = 1, \ldots, n$ to get the information brought by the whole sample. It is asymptotically equivalent to

$$n \frac{(T_2 - T_1)^2}{T(e^{\vartheta(T_2-T_1)} - 1)^2} \int_0^T \frac{e^{-2\vartheta(T_1-t)}\sigma^2_t}{\overline{\sigma}^2_t} dt.$$

We take the inverse of this expression divided by Δ_n^{-1} to get the result, which is a lower bound for the limit variance in the estimation.

2.4.5 Proof of Theorem 2.4

The first assertion has been proved in Section 2.4.4, while proving Theorem 2.3 : the efficient score $\widetilde{\ell}_{\vartheta,\sigma,\overline{\sigma}}$ for ϑ associated to the couple of observations i in the experiment $\mathcal{E}^n$ is

$$\widetilde{\ell}_{\overline{\sigma}}(\vartheta)^i = \frac{(\Delta^n_i X^2 - \Delta^n_i X^1)(\Delta^n_i X^2 - e^{-\vartheta(T_2-T_1)}\Delta^n_i X^1)e^{-\vartheta(T_2-T_1)}(T_2 - T_1)}{(1 - e^{-\vartheta(T_2-T_1)})^3 \int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}^2_t dt},$$

and the notation $\widetilde{\ell}_{\overline{\sigma}}(\vartheta)^i$ emphasizes that σ does not appear in this function.

Now we turn to the core of the proof. As usual, we assume that the drift processes are 0 and that σ and $\overline{\sigma}$ are bounded by some constant $M_\Sigma > 0$, removing those assumptions afterwards.
In the proof, for ease of notation we will write t_i instead of $i\Delta_n$.

First step

Let ϑ_n be a deterministic sequence such that $\sqrt{n}(\vartheta_n - \vartheta) = O(1)$. First we show that we may substitute $\widetilde{\ell}^i_{\vartheta_n,\sigma,\overline{\sigma}}$ by $\widetilde{\ell}\big(\vartheta_n, \widehat{\overline{\sigma}}^2_n\big)^i$ in the estimator, which means that we can plug nonparametric estimators into the efficient score function.
More precisely, what we would like to show is that $\Delta_n^{1/2} \sum_{i\in\mathcal{I}_n} \big(\widetilde{\ell}^i_{\vartheta_n,\sigma,\overline{\sigma}} - \widetilde{\ell}\big(\vartheta_n, \widehat{\overline{\sigma}}^2_n\big)^i\big) \to 0$ in probability, as $n \to \infty$. This amounts to show that

$$\Delta_n^{1/2} \sum_{i\in\mathcal{I}_n} (\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta_n(T_2-T_1)} \Delta_i^n X^1)\Big(\frac{1}{\int_{t_{i-1}}^{t_i} \overline{\sigma}_t^2 dt} - \frac{1}{\Delta_n \widehat{\overline{\sigma}}^2_{n,t_{i-1}}}\Big) \tag{2.9}$$

converges to 0 in probability. We rewrite it as

$$\begin{aligned}
&\Delta_n^{1/2} \sum_{i\in\mathcal{I}_n} (\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta_n(T_2-T_1)} \Delta_i^n X^1)\Big(\frac{1}{\int_{t_{i-1}}^{t_i} \overline{\sigma}_t^2 dt} - \frac{1}{\Delta_n \overline{\sigma}^2_{t_{i-1}}}\Big) \\
&+ \Delta_n^{1/2} \sum_{i\in\mathcal{I}_n} (\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta_n(T_2-T_1)} \Delta_i^n X^1)\frac{1}{\Delta_n}\Big(\frac{1}{\overline{\sigma}^2_{t_{i-1}}} - \frac{1}{\widehat{\overline{\sigma}}^2_{n,t_{i-1}}}\Big) \\
=&S_n' + S_n''
\end{aligned}$$

where

$$S_n' = \Delta_n^{-1/2} \sum_{i\in\mathcal{I}_n} (\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta_n(T_2-T_1)} \Delta_i^n X^1)\frac{\Delta_n \overline{\sigma}^2_{t_{i-1}} - \int_{t_{i-1}}^{t_i} \overline{\sigma}_t^2 dt}{\overline{\sigma}^2_{t_{i-1}} \int_{t_{i-1}}^{t_i} \overline{\sigma}_t^2 dt}$$

and

$$S_n'' = \Delta_n^{-1/2} \sum_{i\in\mathcal{I}_n} (\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta_n(T_2-T_1)} \Delta_i^n X^1)\frac{\widehat{\overline{\sigma}}^2_{t_{i-1}} - \overline{\sigma}^2_{t_{i-1}}}{\overline{\sigma}^2_{t_{i-1}} \widehat{\overline{\sigma}}^2_{t_{i-1}}}.$$

To care for S_n', we have that

$$\mathbb{E}(|S_n'|) \le \sum_{i\in\mathcal{I}_n} \mathbb{E}\Big(\Big|\Delta_n^{-1/2}(\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta_n(T_2-T_1)} \Delta_i^n X^1)\frac{\Delta_n \overline{\sigma}^2_{t_{i-1}} - \int_{t_{i-1}}^{t_i} \overline{\sigma}_t^2 dt}{\overline{\sigma}^2_{t_{i-1}} \int_{t_{i-1}}^{t_i} \overline{\sigma}_t^2 dt}\Big|\Big),$$

and for $i \in \mathcal{I}_n$,

$$\begin{aligned}
&\mathbb{E}\Big(\Big|\Delta_n^{-1/2}(\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta_n(T_2-T_1)} \Delta_i^n X^1)\frac{\Delta_n \overline{\sigma}^2_{t_{i-1}} - \int_{t_{i-1}}^{t_i} \overline{\sigma}_t^2 dt}{\overline{\sigma}^2_{t_{i-1}} \int_{t_{i-1}}^{t_i} \overline{\sigma}_t^2 dt}\Big|\Big) \\
\le&\frac{\Delta_n^{-3/2}}{\tilde{c}^4}\sqrt{\mathbb{E}\Big(\big|(\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta_n(T_2-T_1)} \Delta_i^n X^1)\big|^2\Big)\mathbb{E}\Big(\Big|\Delta_n \overline{\sigma}^2_{t_{i-1}} - \int_{t_{i-1}}^{t_i} \overline{\sigma}_t^2 dt\Big|^2\Big)}
\end{aligned}$$

using Cauchy-Schwarz and the fact that $\mathbb{P}((\sigma, \overline{\sigma}) \in \Sigma(c, \tilde{c})) = 1$. Using again Cauchy-Schwarz and then BDG inequality, we have that $\mathbb{E}\Big(\big|(\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta_n(T_2-T_1)} \Delta_i^n X^1)\big|^2\Big)$

has order Δ_n^2. Also,

$$\begin{aligned}
&\mathbb{E}\Big(\Big|\Delta_n^{-1/2}(\Delta_i^n X^2-\Delta_i^n X^1)(\Delta_i^n X^2-e^{-\vartheta_n(T_2-T_1)}\Delta_i^n X^1)\frac{\Delta_n\overline{\sigma}_{t_{i-1}}^2-\int_{t_{i-1}}^{t_i}\overline{\sigma}_t^2dt}{\overline{\sigma}_{t_{i-1}}^2\int_{t_{i-1}}^{t_i}\overline{\sigma}_t^2dt}\Big|\Big)\\
\lesssim&\Delta_n^{-1/2}\sqrt{\mathbb{E}\Big(\Big|\Delta_n\overline{\sigma}_{t_{i-1}}^2-\int_{t_{i-1}}^{t_i}\overline{\sigma}_t^2dt\Big|^2\Big)}
\end{aligned}$$

and

$$\sqrt{\mathbb{E}\Big(\Big|\Delta_n\overline{\sigma}_{t_{i-1}}^2-\int_{t_{i-1}}^{t_i}\overline{\sigma}_t^2dt\Big|^2\Big)}=\sqrt{\mathbb{E}\Big(\Big|\int_{t_{i-1}}^{t_i}\overline{\sigma}_{t_{i-1}}^2-\overline{\sigma}_t^2dt\Big|^2\Big)},$$

so that, using Jensen inequality,

$$\sqrt{\mathbb{E}\Big(\Big|\Delta_n\overline{\sigma}_{t_{i-1}}^2-\int_{t_{i-1}}^{t_i}\overline{\sigma}_t^2dt\Big|^2\Big)}\le\sqrt{\Delta_n\mathbb{E}\Big(\int_{t_{i-1}}^{t_i}|\overline{\sigma}_{t_{i-1}}^2-\overline{\sigma}_t^2|^2dt\Big)}\le c(\Delta_n)^{1+\alpha}.$$

Finally,

$$\mathbb{E}(|S_n'|)\lesssim\sum_{i\in\mathcal{I}_n}\Delta_n^{-1/2}\Delta_n^{1+\alpha}\lesssim\Delta_n^{\alpha-1/2}.$$

Because $\alpha>1/2$, we conclude that S_n' converges to 0 in L^1 and thus in probability.

Now we look at the term S_n'' : because the kernel used for nonparametric estimation has its support included in $(0,+\infty)$, each $\widehat{\overline{\sigma}}_{n,t_{i-1}}^2$ is $\mathcal{F}_{i-1}$-measurable, and

$$\begin{aligned}
&\mathbb{E}\Big(\Delta_n^{-1/2}(\Delta_i^n X^2-\Delta_i^n X^1)(\Delta_i^n X^2-e^{-\vartheta_n(T_2-T_1)}\Delta_i^n X^1)\frac{\widehat{\overline{\sigma}}_{n,t_{i-1}}^2-\overline{\sigma}_{t_{i-1}}^2}{\overline{\sigma}_{t_{i-1}}^2\widehat{\overline{\sigma}}_{n,t_{i-1}}^2}\Big|\mathcal{F}_{i-1}\Big)\\
=&\Delta_n^{-1/2}\mathbb{E}\big((\Delta_i^n X^2-\Delta_i^n X^1)(\Delta_i^n X^2-e^{-\vartheta_n(T_2-T_1)}\Delta_i^n X^1)\big|\mathcal{F}_{i-1}\big)\frac{\widehat{\overline{\sigma}}_{n,t_{i-1}}^2-\overline{\sigma}_{t_{i-1}}^2}{\overline{\sigma}_{t_{i-1}}^2\widehat{\overline{\sigma}}_{n,t_{i-1}}^2}\\
=&\Delta_n^{-1/2}\big(e^{-\vartheta(T_2-T_1)}-e^{-\vartheta_n(T_2-T_1)}\big)\big(e^{-\vartheta(T_2-T_1)}-1\big)\chi_i^n,
\end{aligned}$$

where

$$\chi_i^n=\mathbb{E}\Big(\int_{t_{i-1}}^{t_i}e^{-2\vartheta(T_1-t)}\sigma_t^2dt\Big|\mathcal{F}_{i-1}\Big)\frac{\widehat{\overline{\sigma}}_{n,t_{i-1}}^2-\overline{\sigma}_{t_{i-1}}^2}{\overline{\sigma}_{t_{i-1}}^2\widehat{\overline{\sigma}}_{n,t_{i-1}}^2}.$$

Now,

$$\mathbb{E}\big(|\chi_i^n|\big|\mathcal{F}_{i-1}\big)\le\Delta_n M_\Sigma^2\frac{\sup_{i\in\mathcal{I}_n}\big|\widehat{\overline{\sigma}}_{n,t_{i-1}}^2-\overline{\sigma}_{t_{i-1}}^2\big|}{\tilde{c}^4},$$

so that $\sum_{i=1}^n\mathbb{E}\big(|\chi_i^n|\big|\mathcal{F}_{i-1}\big)\le\frac{M_\Sigma^2}{\tilde{c}^4}\sup_{i\in\mathcal{I}_n}\big|\widehat{\overline{\sigma}}_{n,t_{i-1}}^2-\overline{\sigma}_{t_{i-1}}^2\big|$, which converges to 0 in probability by Theorem 2.2. As the sequence $\Delta_n^{-1/2}\big(e^{-\vartheta(T_2-T_1)}-e^{-\vartheta_n(T_2-T_1)}\big)$ is tight, we use Lemma 3.4

in [49] applied to variables χ_i^n to conclude that for all $t \in [0,T]$,

$$\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}\Big(\Delta_n^{-1/2}(\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta_n(T_2-T_1)}\Delta_i^n X^1)\frac{\widehat{\overline{\sigma}}^2_{n,t_{i-1}} - \overline{\sigma}^2_{t_{i-1}}}{\overline{\sigma}^2_{t_{i-1}}\widehat{\overline{\sigma}}^2_{n,t_{i-1}}}\Big|\mathcal{F}_{i-1}\Big) \overset{u.c.p.}{\to} 0, \tag{2.10}$$

where $X_t^n \overset{u.c.p.}{\to} X_t$ means "convergence in probability, locally uniformly in time", that is $\sup_{s\le t}|X_s^n - X_s| \to 0$ in probability for all t; this is the definition from Section 3.1 in [49]. Moreover,

$$\begin{aligned}
&\mathbb{E}\Big(\Big(\Delta_n^{-1/2}(\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta_n(T_2-T_1)}\Delta_i^n X^1)\frac{\widehat{\overline{\sigma}}^2_{n,t_{i-1}} - \overline{\sigma}^2_{t_{i-1}}}{\overline{\sigma}^2_{t_{i-1}}\widehat{\overline{\sigma}}^2_{n,t_{i-1}}}\Big)^2\Big|\mathcal{F}_{i-1}\Big)\\
=&\Delta_n^{-1}\mathbb{E}\big(\big((\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta_n(T_2-T_1)}\Delta_i^n X^1)\big)^2\big|\mathcal{F}_{i-1}\big)\Big(\frac{\widehat{\overline{\sigma}}^2_{n,t_{i-1}} - \overline{\sigma}^2_{t_{i-1}}}{\overline{\sigma}^2_{t_{i-1}}\widehat{\overline{\sigma}}^2_{n,t_{i-1}}}\Big)^2\\
\le&\Delta_n^{-1}\mathbb{E}\big(\big((\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta_n(T_2-T_1)}\Delta_i^n X^1)\big)^2\big|\mathcal{F}_{i-1}\big)\frac{\big(\sup_{i\in\mathcal{I}_n}\big|\widehat{\overline{\sigma}}^2_{n,t_{i-1}} - \overline{\sigma}^2_{t_{i-1}}\big|\big)^2}{\tilde{c}^8}.
\end{aligned}$$

As

$$\mathbb{E}\big(\big((\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta_n(T_2-T_1)}\Delta_i^n X^1)\big)^2\big|\mathcal{F}_{i-1}\big)$$

is of order Δ_n^2 and $\sup_i\big|\widehat{\overline{\sigma}}^2_{n,t_{i-1}} - \overline{\sigma}^2_{t_{i-1}}\big|$ is of order $\Delta_n^{\alpha/(2\alpha+1)}$, there remains that

$$\sum_{i\in\mathcal{I}_n}\mathbb{E}\Big(\Big(\Delta_n^{-1/2}(\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta_n(T_2-T_1)}\Delta_i^n X^1)\frac{\widehat{\overline{\sigma}}^2_{n,t_{i-1}} - \overline{\sigma}^2_{t_{i-1}}}{\overline{\sigma}^2_{t_{i-1}}\widehat{\overline{\sigma}}^2_{n,t_{i-1}}}\Big)^2\Big|\mathcal{F}_{i-1}\Big) \lesssim \Delta_n^{\alpha/(2\alpha+1)}$$

which converges to 0 in probability. With this result and (2.10), by Lemma 3.4 in [49], we conclude that S_n'' converges to 0 in probability, which gives the expected result.

Second step

As $\vartheta \mapsto \widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}}$ is regular enough, we can write a Taylor expansion of $\widetilde{\ell}^i_{\vartheta_n,\sigma,\overline{\sigma}}$ at ϑ, and establish that

$$\sqrt{|\mathcal{I}_n|}\Big(\frac{1}{|\mathcal{I}_n|}\sum_{i\in\mathcal{I}_n}\big(\widetilde{\ell}^i_{\vartheta_n,\sigma,\overline{\sigma}} - \widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}}\big) + \frac{1}{|\mathcal{I}_n|}\sum_{i\in\mathcal{I}_n}\widetilde{I}_{i,\vartheta,\sigma,\overline{\sigma}}(\vartheta_n - \vartheta)\Big)$$

converges to 0 in probability, which amounts to say that

$$\Delta_n^{1/2}\Big(\sum_{i\in\mathcal{I}_n}\big(\widetilde{\ell}^i_{\vartheta_n,\sigma,\overline{\sigma}} - \widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}}\big) + \sum_{i\in\mathcal{I}_n}\widetilde{I}_{i,\vartheta,\sigma,\overline{\sigma}}(\vartheta_n - \vartheta)\Big)$$

converges to 0 in probability, because $|\mathcal{I}_n| \sim n(1-h_n) \sim n$.
We introduce the notation $\widetilde{I}_{\vartheta,\sigma,\overline{\sigma}}$ for

$$\mathbb{P}-\lim_{n\to+\infty}\Big[\Delta_n\sum_{i\in\mathcal{I}_n}\widetilde{I}_{i,\vartheta,\sigma,\overline{\sigma}}\Big] = \frac{(T_2-T_1)^2}{(e^{\vartheta T_2} - e^{\vartheta T_1})^2}\int_0^T \frac{e^{2\vartheta t}\sigma_t^2}{\overline{\sigma}_t^2}dt$$

which is equivalent to the information brought by the whole sample divided by the number of observations. We combine this with the fact that (2.9) converges to 0 in probability to get that

$$\Delta_n^{1/2}\Big(\sum_{i\in\mathcal{I}_n}\big(\widetilde{\ell}\big(\vartheta_n,\widehat{\overline{\sigma}}_n^2\big)^i-\widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}}\big)+\Delta_n^{-1}\widetilde{I}_{\vartheta,\sigma,\overline{\sigma}}(\vartheta_n-\vartheta)\Big) \tag{2.11}$$

converges to 0 in probability, which remains true if we replace the deterministic sequence ϑ_n by the discretized version of $\hat{\vartheta}_{2,n}$ (see the proof of Theorem 5.48 in [73] for an argument).

The next step is to prove that

$$\Delta_n\sum_{i\in\mathcal{I}_n}\Big(\widetilde{\ell}\big(\hat{\vartheta}_{2,n},\widehat{\overline{\sigma}}_n^2\big)^i\Big)^2\to\widetilde{I}_{\vartheta,\sigma,\overline{\sigma}} \tag{2.12}$$

in probability. To do so, we observe that $\Delta_n\sum_{i\in\mathcal{I}_n}\Big(\widetilde{\ell}\big(\vartheta_n,\widehat{\overline{\sigma}}_n^2\big)^i\Big)^2-\Delta_n\sum_{i\in\mathcal{I}_n}\big(\widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}}\big)^2$ can be decomposed as $(T_2-T_1)\sum_{i\in\mathcal{I}_n}\mathcal{T}^1_{i,n}+\mathcal{T}^2_{i,n}+\mathcal{T}^3_{i,n}$, where

$$\begin{aligned}\mathcal{T}^1_{i,n}=\Delta_n\frac{(\Delta_i^nX^2-\Delta_i^nX^1)^2}{\big(\Delta_n\widehat{\overline{\sigma}}^2_{n,t_{i-1}}\big)^2}\Big[&\frac{(\Delta_i^nX^2-e^{-\vartheta_n(T_2-T_1)}\Delta_i^nX^1)^2e^{-2\vartheta_n(T_2-T_1)}}{(1-e^{-\vartheta_n(T_2-T_1)})^6}\\&-\frac{(\Delta_i^nX^2-e^{-\vartheta(T_2-T_1)}\Delta_i^nX^1)^2e^{-2\vartheta(T_2-T_1)}}{(1-e^{-\vartheta(T_2-T_1)})^6}\Big],\end{aligned}$$

while $\mathcal{T}^2_{i,n}$ and $\mathcal{T}^3_{i,n}$ are respectively

$$\Delta_n(\Delta_i^nX^2-\Delta_i^nX^1)^2\frac{(\Delta_i^nX^2-e^{-\vartheta(T_2-T_1)}\Delta_i^nX^1)^2e^{-2\vartheta(T_2-T_1)}}{(1-e^{-\vartheta(T_2-T_1)})^6}\Big(\frac{1}{\big(\Delta_n\widehat{\overline{\sigma}}^2_{n,t_{i-1}}\big)^2}-\frac{1}{\big(\Delta_n\overline{\sigma}^2_{t_{i-1}}\big)^2}\Big)$$

and

$$\Delta_n(\Delta_i^nX^2-\Delta_i^nX^1)^2\frac{(\Delta_i^nX^2-e^{-\vartheta(T_2-T_1)}\Delta_i^nX^1)^2e^{-2\vartheta(T_2-T_1)}}{(1-e^{-\vartheta(T_2-T_1)})^6}\Big(\frac{1}{\big(\Delta_n\overline{\sigma}^2_{t_{i-1}}\big)^2}-\frac{1}{\big(\int_{t_{i-1}}^{t_i}\overline{\sigma}_t^2dt\big)^2}\Big).$$

First,

$$\begin{aligned}&\mathbb{E}\Big(\Big|(\Delta_i^nX^2-\Delta_i^nX^1)^2\frac{(\Delta_i^nX^2-e^{-\vartheta(T_2-T_1)}\Delta_i^nX^1)^2}{(1-e^{-\vartheta(T_2-T_1)})^6}\Big(\frac{1}{\big(\Delta_n\widehat{\overline{\sigma}}^2_{n,t_{i-1}}\big)^2}-\frac{1}{\big(\Delta_n\overline{\sigma}^2_{t_{i-1}}\big)^2}\Big)\Big||\mathcal{F}_{i-1}\Big)\\=&\mathbb{E}\Big(\Big|(\Delta_i^nX^2-\Delta_i^nX^1)^2\frac{(\Delta_i^nX^2-e^{-\vartheta(T_2-T_1)}\Delta_i^nX^1)^2}{(1-e^{-\vartheta(T_2-T_1)})^6}\frac{\overline{\sigma}^4_{t_{i-1}}-\widehat{\overline{\sigma}}^4_{n,t_{i-1}}}{\Delta_n^2\widehat{\overline{\sigma}}^4_{n,t_{i-1}}\overline{\sigma}^4_{t_{i-1}}}\Big||\mathcal{F}_{i-1}\Big),\end{aligned}$$

and as $\overline{\sigma}_{t_{i-1}}$ and $\widehat{\overline{\sigma}}_{n,t_{i-1}}$ are $\mathcal{F}_i$-measurable,

$$\begin{aligned}&\mathbb{E}\Big(\Big|(\Delta_i^nX^2-\Delta_i^nX^1)^2\frac{(\Delta_i^nX^2-e^{-\vartheta(T_2-T_1)}\Delta_i^nX^1)^2}{(1-e^{-\vartheta(T_2-T_1)})^6}\Big(\frac{1}{\big(\Delta_n\widehat{\overline{\sigma}}^2_{n,t_{i-1}}\big)^2}-\frac{1}{\big(\Delta_n\overline{\sigma}^2_{t_{i-1}}\big)^2}\Big)\Big||\mathcal{F}_{i-1}\Big)\\=&\mathbb{E}\Big(\Big|(\Delta_i^nX^2-\Delta_i^nX^1)^2\frac{(\Delta_i^nX^2-e^{-\vartheta(T_2-T_1)}\Delta_i^nX^1)^2}{(1-e^{-\vartheta(T_2-T_1)})^6}\Big||\mathcal{F}_{i-1}\Big)\Big|\frac{\overline{\sigma}^4_{t_{i-1}}-\widehat{\overline{\sigma}}^4_{n,t_{i-1}}}{\Delta_n^2\widehat{\overline{\sigma}}^4_{n,t_{i-1}}\overline{\sigma}^4_{t_{i-1}}}\Big|.\end{aligned}$$

As

$$\overline{\sigma}^4_{t_{i-1}} - \widehat{\overline{\sigma}}^4_{n,t_{i-1}} = \big(\overline{\sigma}^2_{t_{i-1}} - \widehat{\overline{\sigma}}^2_{n,t_{i-1}}\big)\big(\overline{\sigma}^2_{t_{i-1}} + \widehat{\overline{\sigma}}^2_{n,t_{i-1}}\big) = \big(\overline{\sigma}^2_{t_{i-1}} - \widehat{\overline{\sigma}}^2_{n,t_{i-1}}\big)\big(2\overline{\sigma}^2_{t_{i-1}} + \widehat{\overline{\sigma}}^2_{n,t_{i-1}} - \overline{\sigma}^2_{t_{i-1}}\big),$$

we have

$$\big|\overline{\sigma}^4_{t_{i-1}} - \widehat{\overline{\sigma}}^4_{n,t_{i-1}}\big| \le \Big(\sup_{i\in\mathcal{I}_n}\big|\overline{\sigma}^2_{t_{i-1}} - \widehat{\overline{\sigma}}^2_{n,t_{i-1}}\big|\Big)\Big(2M_\Sigma^2 + \sup_{i\in\mathcal{I}_n}\big|\overline{\sigma}^2_{t_{i-1}} - \widehat{\overline{\sigma}}^2_{n,t_{i-1}}\big|\Big),$$

so that

$$\mathbb{E}\big(|\mathcal{T}^2_{i,n}|\big|\mathcal{F}_{i-1}\big) \lesssim \Big(\sup_{i\in\mathcal{I}_n}\big|\overline{\sigma}^2_{t_{i-1}} - \widehat{\overline{\sigma}}^2_{n,t_{i-1}}\big|\Big)\Big(2M_\Sigma^2 + \sup_{i\in\mathcal{I}_n}\big|\overline{\sigma}^2_{t_{i-1}} - \widehat{\overline{\sigma}}^2_{n,t_{i-1}}\big|\Big).$$

We invoke Theorem 2.2 which asserts that $\sup_i\big|\overline{\sigma}^2_{t_{i-1}} - \widehat{\overline{\sigma}}^2_{n,t_{i-1}}\big|$ converges to 0 in probability, and Lemma 3.4 in [49] to prove that $\sum_{i\in\mathcal{I}_n}\mathcal{T}^2_{i,n}$ converges to 0 in probability. Then,

$$\begin{aligned}
&\mathbb{E}\Big(\Big|\Delta_n(\Delta_i^n X^2 - \Delta_i^n X^1)^2(\Delta_i^n X^2 - e^{-\vartheta(T_2-T_1)}\Delta_i^n X^1)^2\Big(\frac{1}{\big(\Delta_n\overline{\sigma}^2_{t_{i-1}}\big)^2} - \frac{1}{\big(\int_{t_{i-1}}^{t_i}\overline{\sigma}^2_t dt\big)^2}\Big)\Big|\Big)\\
=&\mathbb{E}\Big(\Big|\Delta_n(\Delta_i^n X^2 - \Delta_i^n X^1)^2(\Delta_i^n X^2 - e^{-\vartheta(T_2-T_1)}\Delta_i^n X^1)^2\frac{\big(\int_{t_{i-1}}^{t_i}\overline{\sigma}^2_t dt\big)^2 - \Delta_n^2\overline{\sigma}^4_{t_{i-1}}}{\Delta_n^2\overline{\sigma}^4_{t_{i-1}}\big(\int_{t_{i-1}}^{t_i}\overline{\sigma}^2_t dt\big)^2}\Big|\Big)\\
\le&\Delta_n\sqrt{\mathbb{E}\Big(\Big|(\Delta_i^n X^2 - \Delta_i^n X^1)^4(\Delta_i^n X^2 - e^{-\vartheta(T_2-T_1)}\Delta_i^n X^1)^4\Big|\Big)\mathbb{E}\Big(\Big|\frac{\Big(\big(\int_{t_{i-1}}^{t_i}\overline{\sigma}^2_t dt\big)^2 - \Delta_n^2\overline{\sigma}^4_{t_{i-1}}\Big)^2}{\Delta_n^4\overline{\sigma}^8_{t_{i-1}}\big(\int_{t_{i-1}}^{t_i}\overline{\sigma}^2_t dt\big)^4}\Big|\Big)},
\end{aligned}$$

using Cauchy-Schwarz inequality.
We have that $\mathbb{E}\big(\big|(\Delta_i^n X^2 - \Delta_i^n X^1)^4(\Delta_i^n X^2 - e^{-\vartheta(T_2-T_1)}\Delta_i^n X^1)^4\big|\big)$ is of order Δ_n^4, and

$$\begin{aligned}
&\mathbb{E}\Big(\Big|\frac{\Big(\big(\int_{t_{i-1}}^{t_i}\overline{\sigma}^2_t dt\big)^2 - \Delta_n^2\overline{\sigma}^4_{t_{i-1}}\Big)^2}{\Delta_n^4\overline{\sigma}^8_{t_{i-1}}\big(\int_{t_{i-1}}^{t_i}\overline{\sigma}^2_t dt\big)^4}\Big|\Big)\\
\le&\frac{1}{\Delta_n^8\tilde{c}^{16}}\mathbb{E}\Big(\Big(\Big(\int_{t_{i-1}}^{t_i}\overline{\sigma}^2_t dt - \Delta_n\overline{\sigma}^2_{t_{i-1}}\Big)\Big(\int_{t_{i-1}}^{t_i}\overline{\sigma}^2_t dt + \Delta_n\overline{\sigma}^2_{t_{i-1}}\Big)\Big)^2\Big)\\
\le&\frac{\big(2\Delta_n M_\Sigma\big)^2}{\Delta_n^8\tilde{c}^{16}}\mathbb{E}\Big(\Big(\int_{t_{i-1}}^{t_i}\overline{\sigma}^2_t dt - \Delta_n\overline{\sigma}^2_{t_{i-1}}\Big)^2\Big)\\
\le&\frac{4M_\Sigma^2}{\Delta_n^6\tilde{c}^{16}}\Delta_n\int_{t_{i-1}}^{t_i}\mathbb{E}\big((\overline{\sigma}^2_t - \overline{\sigma}^2_{t_{i-1}})^2\big)dt\\
\lesssim&\Delta_n^{2\alpha-4},
\end{aligned}$$

using Jensen inequality and Assumption 2.2. We thus get that $\mathbb{E}(\sum_{i\in\mathcal{I}_n}\mathcal{T}^3_{i,n}) \lesssim \Delta_n^\alpha$, which goes to 0 as $n\to\infty$, and also that $\sum_{i\in\mathcal{I}_n}\mathcal{T}^3_{i,n}$ converges to 0 in probability. To care for $\mathcal{T}^1_{i,n}$, by Taylor theorem, there exists some $\tilde{\vartheta}_n \in [\vartheta, \vartheta_n]$ such that

$$\begin{aligned}
&\frac{(\Delta_i^n X^2 - e^{-\vartheta_n(T_2-T_1)}\Delta_i^n X^1)^2 e^{-2\vartheta_n(T_2-T_1)}}{(1-e^{-\vartheta_n(T_2-T_1)})^6} - \frac{(\Delta_i^n X^2 - e^{-\vartheta(T_2-T_1)}\Delta_i^n X^1)^2 e^{-2\vartheta(T_2-T_1)}}{(1-e^{-\vartheta(T_2-T_1)})^6}\\
=&(\vartheta_n - \vartheta)\frac{d}{d\vartheta}\Big|_{\vartheta=\tilde{\vartheta}_n}\Big(\frac{(\Delta_i^n X^2 - e^{-\vartheta(T_2-T_1)}\Delta_i^n X^1)^2 e^{-2\vartheta(T_2-T_1)}}{(1-e^{-\vartheta(T_2-T_1)})^6}\Big)\\
=&(\vartheta_n - \vartheta)\frac{(T_2-T_1)(\Delta_i^n X^2 - e^{-\tilde{\vartheta}_n(T_2-T_1)}\Delta_i^n X^1)e^{-\tilde{\vartheta}_n(T_2-T_1)}}{(1-e^{-\tilde{\vartheta}_n(T_2-T_1)})^7}\\
&\times\big(3e^{-\tilde{\vartheta}_n(T_2-T_1)}(1+e^{-\tilde{\vartheta}_n(T_2-T_1)})\Delta_i^n X^1 - 2(1-e^{-\tilde{\vartheta}_n(T_2-T_1)})\Delta_i^n X^2\big).
\end{aligned}$$

Then, for n high enough, $|\tilde{\vartheta}_n|$ is less than, say, 2ϑ, and thus

$$\begin{aligned}
&\mathbb{E}\Big[(\Delta_i^n X^2 - \Delta_i^n X^1)^2(\Delta_i^n X^2 - e^{-\tilde{\vartheta}_n(T_2-T_1)}\Delta_i^n X^1)\\
&\quad\times\big(3e^{-\tilde{\vartheta}_n(T_2-T_1)}(1+e^{-\tilde{\vartheta}_n(T_2-T_1)})\Delta_i^n X^1 - 2(1-e^{-\tilde{\vartheta}_n(T_2-T_1)})\Delta_i^n X^2\big)\Big]
\end{aligned}$$

is of order Δ_n^2, so that

$$\mathbb{E}(|\mathcal{T}_{i,n}^1|) \lesssim |\vartheta_n - \vartheta|\frac{\Delta_n^{-1}}{\tilde{c}^4}\Delta_n^2,$$

and also,

$$\mathbb{E}\Big(\Big|\sum_{i\in\mathcal{I}_n}\mathcal{T}_{i,n}^1\Big|\Big) \lesssim |\vartheta_n - \vartheta|,$$

which converges to 0 in probability.
This proves that $\Delta_n\sum_{i\in\mathcal{I}_n}\Big(\widetilde{\ell}\big(\vartheta_n, \widehat{\overline{\sigma}}_n^2\big)^i\Big)^2 - \Delta_n\sum_{i\in\mathcal{I}_n}\big(\widetilde{\ell}_{\vartheta,\sigma,\overline{\sigma}}^i\big)^2$ converges to 0 in probability. As we said for (2.11), this result remains true with $\hat{\vartheta}_{2,n}$ in place of ϑ_n, so that

$$\Delta_n\sum_{i\in\mathcal{I}_n}\big(\widetilde{\ell}\big(\hat{\vartheta}_{2,n}, \widehat{\overline{\sigma}}_n^2\big)^i\big)^2 = \Delta_n\sum_{i\in\mathcal{I}_n}\big(\widetilde{\ell}_{\vartheta,\sigma,\overline{\sigma}}^i\big)^2 + o_{\mathbb{P}}(1) = \widetilde{I}_{\vartheta,\sigma,\overline{\sigma}} + o_{\mathbb{P}}(1),$$

which proves (2.12).
We then have

$$\begin{aligned}
\Delta_n^{-1/2}\big(\tilde{\vartheta}_{2,n} - \vartheta\big)\widetilde{I}_{\vartheta,\sigma,\overline{\sigma}} =&\Delta_n^{-1/2}(\hat{\vartheta}_{2,n} - \vartheta)\widetilde{I}_{\vartheta,\sigma,\overline{\sigma}} + \Delta_n^{-1/2}\frac{\Delta_n\widetilde{I}_{\vartheta,\sigma,\overline{\sigma}}\sum_{i\in\mathcal{I}_n}\widetilde{\ell}\big(\hat{\vartheta}_{2,n}, \widehat{\overline{\sigma}}_n^2\big)^i}{\Delta_n\sum_{i\in\mathcal{I}_n}\big(\widetilde{\ell}\big(\hat{\vartheta}_{2,n}, \widehat{\overline{\sigma}}_n^2\big)^i\big)^2}\\
=&\Delta_n^{-1/2}(\hat{\vartheta}_{2,n} - \vartheta)\widetilde{I}_{\vartheta,\sigma,\overline{\sigma}} + \Delta_n^{-1/2}\Delta_n\sum_{i\in\mathcal{I}_n}\widetilde{\ell}\big(\hat{\vartheta}_{2,n}, \widehat{\overline{\sigma}}_n^2\big)^i\\
&+\Delta_n^{-1/2}\Delta_n\widetilde{I}_{\vartheta,\sigma,\overline{\sigma}}\sum_{i\in\mathcal{I}_n}\widetilde{\ell}\big(\hat{\vartheta}_{2,n}, \widehat{\overline{\sigma}}_n^2\big)^i\Big(\frac{1}{\Delta_n\sum_{i\in\mathcal{I}_n}\big(\widetilde{\ell}\big(\hat{\vartheta}_{2,n}, \widehat{\overline{\sigma}}_n^2\big)^i\big)^2} - \frac{1}{\widetilde{I}_{\vartheta,\sigma,\overline{\sigma}}}\Big)\\
=&\Delta_n^{1/2}\sum_{i\in\mathcal{I}_n}\widetilde{\ell}_{\vartheta,\sigma,\overline{\sigma}}^i + o_{\mathbb{P}}(1)
\end{aligned}$$

because of the convergence in probability of (2.11) towards 0 and the convergence (2.12). In view of Lemma 2.4.4, which is stated and proved in the appendix (p. 78), the sum $\Delta_n^{1/2}\big(\widetilde{I}_{\vartheta,\sigma,\overline{\sigma}}\big)^{-1/2}\sum_{i=1}^n \widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}}$ converges stably in law to a random variable defined on $(\widetilde{\Omega},\widetilde{\mathcal{F}},\widetilde{\mathbb{P}})$, which conditionally to $\mathcal{F}$ is Gaussian with variance $\big(\widetilde{I}_{\vartheta,\sigma,\overline{\sigma}}\big)^{-1}$. Now, one technical point remains. As we have

$$\Delta_n^{1/2}\sum_{i=1}^n \widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}} = \Delta_n^{1/2}\sum_{i\in\mathcal{I}_n} \widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}} + \Delta_n^{1/2}\sum_{i=1}^{\lfloor h_n\Delta_n^{-1}\rfloor} \widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}},$$

we shall show below that $\Delta_n^{1/2}\sum_{i=1}^{\lfloor h_n\Delta_n^{-1}\rfloor} \widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}} \to 0$ in probability in order to get the final result, as the two remaining sums will have the same limit in law.
To do so, we write $\Delta_n^{1/2}\widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}} = \frac{(T_2-T_1)e^{-\vartheta(T_2-T_1)}}{(1-e^{-\vartheta(T_2-T_1)})^3}(A_{i,n}+B_{i,n})$, with

$$A_{i,n} = \Delta_n^{1/2}\frac{(\Delta_i^n X^2-\Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta(T_2-T_1)}\Delta_i^n X^1)}{\Delta_n\overline{\sigma}^2_{t_{i-1}}},$$
$$B_{i,n} = \Delta_n^{1/2}(\Delta_i^n X^2-\Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta(T_2-T_1)}\Delta_i^n X^1)\frac{\int_{t_{i-1}}^{t_i}(\overline{\sigma}^2_{t_{i-1}}-\overline{\sigma}^2_t)dt}{\Delta_n\overline{\sigma}^2_{t_{i-1}}\int_{t_{i-1}}^{t_i}\overline{\sigma}^2_t dt}.$$

We have

$$\mathbb{E}\Big(\Big(\sum_{i=1}^{\lfloor h_n\Delta_n^{-1}\rfloor} A_{i,n}\Big)^2\Big) = \sum_{i=1}^{\lfloor h_n\Delta_n^{-1}\rfloor}\mathbb{E}\big(A_{i,n}^2\big)$$

as, for $i<j$, $\mathbb{E}(A_{i,n}A_{j,n}) = \mathbb{E}(A_{i,n}\mathbb{E}(A_{j,n}|\mathcal{F}_{j-1})) = 0$. Moreover,

$$\mathbb{E}\big(A_{i,n}^2\big) \le \frac{\Delta_n^{-1}}{\widetilde{c}^2}\mathbb{E}\big((\Delta_i^n X^2-\Delta_i^n X^1)^2(\Delta_i^n X^2 - e^{-\vartheta(T_2-T_1)}\Delta_i^n X^1)^2\big) \lesssim \Delta_n,$$

so that

$$\mathbb{E}\Big(\Big(\sum_{i=1}^{\lfloor h_n\Delta_n^{-1}\rfloor} A_{i,n}\Big)^2\Big) \lesssim h_n\Delta_n^{-1}\Delta_n = h_n,$$

which converges to 0 as $n\to\infty$. Therefore we have $\sum_{i=1}^{\lfloor h_n\Delta_n^{-1}\rfloor} A_{i,n}\to 0$ in quadratic mean, and thus also in probability.

Then, by Cauchy-Scharz inequality, $\mathbb{E}(|B_{i,n}|)$ is less than the product

$$\frac{\Delta_n^{-3/2}}{\widetilde{c}^4}\sqrt{\mathbb{E}\big((\Delta_i^n X^2-\Delta_i^n X^1)^2(\Delta_i^n X^2 - e^{-\vartheta(T_2-T_1)}\Delta_i^n X^1)^2\big)\mathbb{E}\Big(\Big(\int_{t_{i-1}}^{t_i}(\overline{\sigma}^2_{t_{i-1}}-\overline{\sigma}^2_t)dt\Big)^2\Big)}.$$

As usual, $\mathbb{E}\big((\Delta_i^n X^2-\Delta_i^n X^1)^2(\Delta_i^n X^2 - e^{-\vartheta(T_2-T_1)}\Delta_i^n X^1)^2\big)$ is of order Δ_n^2 and

$$\mathbb{E}\Big(\Big(\int_{t_{i-1}}^{t_i}(\overline{\sigma}^2_{t_{i-1}}-\overline{\sigma}^2_t)dt\Big)^2\Big) \le \Delta_n\int_{t_{i-1}}^{t_i}\mathbb{E}\big((\overline{\sigma}^2_{t_{i-1}}-\overline{\sigma}^2_t)^2\big)dt \lesssim \Delta_n^{2+2\alpha}$$

by Jensen inequality and Assumption 2.2. We therefore get that $\mathbb{E}(|B_{i,n}|) \lesssim \Delta_n^{1/2+\alpha}$, and thus

$$\mathbb{E}\Big(\Big|\sum_{i=1}^{\lfloor h_n\Delta_n^{-1}\rfloor} B_{i,n}\Big|\Big) \leq \sum_{i=1}^{\lfloor h_n\Delta_n^{-1}\rfloor} \mathbb{E}(|B_{i,n}|) \lesssim h_n\Delta_n^{-1}\Delta_n^{1/2+\alpha} = h_n\Delta_n^{\alpha-1/2},$$

which goes to 0 as $n \to \infty$, because $\alpha \geq 1/2$. We conclude that $\sum_{i=1}^{\lfloor h_n\Delta_n^{-1}\rfloor} B_{i,n} \to 0$ in L^1 and thus in probability, and this ends the whole proof.

2.5 Appendices

2.5.1 Proof of Lemma 2.4.1

First part of the lemma

As we assumed that the process Y satisfies, for some $s > 1/2$, $\sup_{t\in[0,T]} t^{-s}\omega(Y)_t < \infty$, we have that Y is continuous in probability on $[0,T]$, and thus uniformly continuous in probability on $[0,T]$. We split the term of interest into three parts:

$$\Delta_n^{-1}\sum_{i=1}^{n}\left(\overline{\Delta}_i^n(Y)\right)^2 - \int_0^T Y_t^2 dt = S_n^1 + S_n^2 + S_n^3,$$

where

$$\begin{aligned} S_n^1 &= \sum_{i=1}^{n}\int_{(i-1)\Delta_n}^{i\Delta_n}(Y_{(i-1)\Delta_n}^2 - Y_t^2)dt, \\ S_n^2 &= \Delta_n^{-1}\sum_{i=1}^{n}\Big(\int_{(i-1)\Delta_n}^{i\Delta_n}(Y_t - Y_{(i-1)\Delta_n})dt\Big)^2, \\ S_n^3 &= 2\sum_{i=1}^{n} Y_{t_{i-1}}\int_{(i-1)\Delta_n}^{i\Delta_n}(Y_t - Y_{(i-1)\Delta_n})dt. \end{aligned}$$

First, fix $\epsilon > 0$. There exists some $\eta > 0$ such that $\mathbb{E}(|Y_t - Y_s|) < \epsilon$ as soon as $|t-s| < \eta$. Moreover, by localization we may assume that there is some constant $M_Y > 0$ such that $|Y_t| \leq M_Y$ for all $t \in [0,T]$. Then

$$\mathbb{E}(|S_n^1|) \leq \sum_{i=1}^{n}\int_{(i-1)\Delta_n}^{i\Delta_n}\mathbb{E}(|Y_{(i-1)\Delta_n}^2 - Y_t^2)|)dt$$

and

$$\mathbb{E}(|Y_{(i-1)\Delta_n}^2 - Y_t^2)|) = \mathbb{E}(|Y_{(i-1)\Delta_n} - Y_t||Y_{(i-1)\Delta_n} + Y_t|) \leq 2M_Y\mathbb{E}(|Y_{(i-1)\Delta_n} - Y_t|),$$

so that if n is high enough, we have $\Delta_n \leq \eta$ and thus,

$$\mathbb{E}(|S_n^1|) \leq \sum_{i=1}^{n}\int_{(i-1)\Delta_n}^{i\Delta_n} 2M_Y\epsilon dt \leq 2TM_Y\epsilon.$$

Thus, S_n^1 converges to 0 in L^1, and therefore in probability. Then,

$$\mathbb{E}(|S_n^2|) \leq \Delta_n^{-1}\sum_{i=1}^{n}\Delta_n\int_{(i-1)\Delta_n}^{i\Delta_n}\mathbb{E}\big((Y_t - Y_{(i-1)\Delta_n})^2\big)dt,$$

and

$$\mathbb{E}\big((Y_t - Y_{(i-1)\Delta_n})^2\big) \leq \mathbb{E}\big(|Y_t - Y_{(i-1)\Delta_n}||Y_t - Y_{(i-1)\Delta_n}|\big) \leq 2M_Y \mathbb{E}\big(|Y_t - Y_{(i-1)\Delta_n}|\big) \leq 2M_Y \epsilon$$

as soon as $\Delta_n \leq \eta$. We thus have $\mathbb{E}(|S_n^2|) \leq 2TM_Y\epsilon$ likewise, which leads to the convervence in probability of S_n^2 towards 0. Finally,

$$\mathbb{E}(|S_n^3|) \leq 2\sum_{i=1}^{n} M_Y \int_{(i-1)\Delta_n}^{i\Delta_n} \mathbb{E}(|Y_t - Y_{(i-1)\Delta_n}|)dt \leq 2TM_Y\epsilon$$

when $\Delta_n \leq \eta$. We have thus proved that $\Delta_n^{-1}\sum_{i=1}^{n}\big(\overline{\Delta}_i^n(Y)\big)^2 - \int_0^T Y_t^2 dt$ converges to 0 in probability, which ends the proof of the first part of the lemma.

Second part of the lemma

We begin by splitting the term of interest as follows:

$$\begin{aligned}
\Delta_n^{-1}\sum_{i=1}^{n}\overline{\Delta}_i^n(Y)\int_{(i-1)\Delta_n}^{i\Delta_n} Z_t dB_t =& \sum_{i=1}^{n}\int_{(i-1)\Delta_n}^{i\Delta_n} Y_t Z_t dB_t \\
&+ \sum_{i=1}^{n}\int_{(i-1)\Delta_n}^{i\Delta_n}(Y_{(i-1)\Delta_n} - Y_t)Z_t dB_t \\
&+ \Delta_n^{-1}\sum_{i=1}^{n}\big(\overline{\Delta}_i^n(Y) - \Delta_n Y_{(i-1)\Delta_n}\big)\int_{(i-1)\Delta_n}^{i\Delta_n} Z_t dB_t \\
=& \int_0^T Y_t Z_t dB_t + S_n^1 + S_n^2
\end{aligned}$$

where

$$S_n^1 = \sum_{i=1}^{n}\int_{(i-1)\Delta_n}^{i\Delta_n}(Y_{(i-1)\Delta_n} - Y_t)Z_t dB_t$$

and

$$S_n^2 = \Delta_n^{-1}\sum_{i=1}^{n}\big(\overline{\Delta}_i^n(Y) - \Delta_n Y_{(i-1)\Delta_n}\big)\int_{(i-1)\Delta_n}^{i\Delta_n} Z_t dB_t.$$

Let us care for the term S_n^1 : for all $i = 1, \ldots, n$, as $\big(\int_{(i-1)\Delta_n}^{t}(Y_{(i-1)\Delta_n} - Y_t)Z_t dB_t\big)_{t\geq(i-1)\Delta_n}$ is a martingale conditionally to $\mathcal{F}_{i-1}$, we have

$$\mathbb{E}\Big(\int_{(i-1)\Delta_n}^{i\Delta_n}(Y_{(i-1)\Delta_n} - Y_t)Z_t dB_t\Big|\mathcal{F}_{i-1}\Big) = 0,$$

and thus, if $1 \leq i < j \leq n$,

$$\begin{aligned}
&\mathbb{E}\Big(\int_{(i-1)\Delta_n}^{i\Delta_n}(Y_{(i-1)\Delta_n}-Y_t)Z_t dB_t \int_{(j-1)\Delta_n}^{j\Delta_n}(Y_{(j-1)\Delta_n}-Y_t)Z_t dB_t\Big)\\
=&\mathbb{E}\Big(\int_{(i-1)\Delta_n}^{i\Delta_n}(Y_{(i-1)\Delta_n}-Y_t)Z_t dB_t \mathbb{E}\Big(\int_{(j-1)\Delta_n}^{j\Delta_n}(Y_{(j-1)\Delta_n}-Y_t)Z_t dB_t\Big|\mathcal{F}_{j-1}\Big)\Big)\\
=&0.
\end{aligned}$$

By localization, we may assume that the càdlàg process Z is such that $|Z_t| < M_Z$ for all $t \in [0,T]$, for some constant $M_Z > 0$. Then, by Tchebychev inequality, for $\varepsilon > 0$,

$$\begin{aligned}
\mathbb{P}(|S_n^1| > \varepsilon) \leq& \frac{1}{\varepsilon^2}\mathbb{E}((S_n^1)^2)\\
=& \frac{1}{\varepsilon^2}\sum_{i=1}^n \mathbb{E}\Big(\Big(\int_{(i-1)\Delta_n}^{i\Delta_n}(Y_{(i-1)\Delta_n}-Y_t)Z_t dB_t\Big)^2\Big)\\
=& \frac{1}{\varepsilon^2}\sum_{i=1}^n \int_{(i-1)\Delta_n}^{i\Delta_n}\mathbb{E}\Big((Y_{(i-1)\Delta_n}-Y_t)^2 Z_t^2\Big)dt\\
\leq& \frac{M_Z^2}{\varepsilon^2}\sum_{i=1}^n \int_{(i-1)\Delta_n}^{i\Delta_n}\mathbb{E}\Big((Y_{(i-1)\Delta_n}-Y_t)^2\Big)dt\\
\leq& \frac{M_Z^2}{\varepsilon^2}\sum_{i=1}^n \int_{(i-1)\Delta_n}^{i\Delta_n} 2M_Y\epsilon dt\\
\leq& 2TM_Y\frac{M_Z^2}{\varepsilon^2}\epsilon,
\end{aligned}$$

as soon as $\Delta_n < \eta$. We have used successively the fact that the cross expectations are zero, Itô isometry and bounds for Z and for $Y_{(i-1)\Delta_n} - Y_t$, in probability. As the choice of ϵ was arbitrary, we have that S_n^1 converges to 0 in probability.

Now, consider the term S_n^2. By Cauchy-Schwarz inequality,

$$S_n^2 \leq \Big(\sum_{i=1}^n\Big(\Delta_n^{-1}\int_{t_{i-1}}^{t_i}(Y_t - Y_{t_{i-1}})dt\Big)^2\Big)^{1/2}\Big(\sum_{i=1}^n\Big(\int_{t_{i-1}}^{t_i} Z_t dB_t\Big)^2\Big)^{1/2}.$$

The second term in the product is $O_{\mathbb{P}}(1)$. We are treating the other one in the context of projection on the Haar basis (see the book [39] for a presentation). Let $\varphi(x) = \mathbf{1}_{[0,1)}$ be the Haar scaling function, and $\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k) = 2^{j/2}\mathbf{1}_{[k2^{-j},(k+1)2^{-j})}$ for any $j \geq 0, k \in \mathbb{Z}$, and let $n = 2^j$:

$$\sum_{i=1}^{n} \Big(\Delta_n^{-1} \int_{t_{i-1}}^{t_i} (Y_t - Y_{t_{i-1}})dt\Big)^2 = \frac{1}{T^2} \sum_{k=1}^{2^j} \Big(2^j \int_{t_{k-1}}^{t_k} (Y_t - Y_{t_{k-1}})dt\Big)^2$$
$$= \sum_{k=1}^{2^j} \Big(2^j \int_{\frac{k-1}{2^j}}^{\frac{k}{2^j}} (Y_{uT} - Y_{\frac{k-1}{2^j}T})du\Big)^2$$
$$= \sum_{k=0}^{2^j-1} \Big(2^{j/2} \int_{\frac{k}{2^j}}^{\frac{k+1}{2^j}} \varphi_{j,k}(u) Y_{uT} du - Y_{\frac{k}{2^j}T}\Big)^2$$
$$= \sum_{k=0}^{2^j-1} \Big(\big(\int \varphi_{j,k} Y_{\cdot T}\big)\varphi_{j,k}\big(\frac{k}{2^j}\big) - Y_{\frac{k}{2^j}T}\Big)^2$$
$$= \sum_{k=0}^{2^j-1} \Big(\big(K_j(Y_{\cdot T})\big(\frac{k}{2^j}\big) - Y_{\frac{k}{2^j}T}\Big)^2$$

where $K_j(f) = \sum_{k=0}^{2^j-1} \big(\int \varphi_{j,k} f\big)\varphi_{j,k}$ is an orthogonal projection operator. Then, for j being high enough, there exists some constant C such that

$$\mathbb{E}\Big(\sum_{k=0}^{2^j-1} \Big(K_j(Y_{\cdot T})\big(\frac{k}{2^j}\big) - Y_{\frac{k}{2^j}T}\Big)^2\Big) \leq C2^j \mathbb{E}\Big(\int_0^1 (K_j(Y_{\cdot T})(u) - Y_{uT})^2 du\Big).$$

Now, for $j \geq 1$,

$$\mathbb{E}\Big(\sum_{i=1}^{n} \Big(\Delta_n^{-1} \int_{t_{i-1}}^{t_i} (Y_t - Y_{t_{i-1}})dt\Big)^2\Big) \leq C2^j \mathbb{E}\Big(\int_0^1 (K_j(Y_{\cdot T})(u) - Y_{uT})^2 du\Big)$$
$$= C2^j \mathbb{E}\Big(\int_0^1 \Big(\int_0^1 \varphi_{j,k}(y)\varphi_{j,k}(u)(Y_{yT} - Y_{uT})dy\Big)^2 du\Big)$$
$$\leq C2^j \mathbb{E}\Big(\int_0^1 \Big(\int_0^1 2^j \mathbf{1}_{|u-y|\in[0,2^{-j})}(Y_{yT} - Y_{uT})dy\Big)^2 du\Big)$$

as the product $\varphi_{j,k}(y)\varphi_{j,k}(u)$ is equal to $2^j \mathbf{1}_{[k2^{-j},(k+1)2^{-j})}(y)\mathbf{1}_{[k2^{-j},(k+1)2^{-j})}(u)$, and as soon as it is non-zero, we have that $k2^{-j} \leq u, y < (k+1)2^{-j}$. Then, by a change of variable and the generalized Minkowski inequality,

$$\begin{aligned}
&\mathbb{E}\Big(\sum_{i=1}^{n}\Big(\Delta_n^{-1}\int_{t_{i-1}}^{t_i}(Y_t-Y_{t_{i-1}})dt\Big)^2\Big)\\
\leq& C2^j\mathbb{E}\Big(\int_0^1\Big(\int_{-2^j u}^{2^j(1-u)}2^j\mathbf{1}_{|x|\in[0,1)}(Y_{uT+2^{-j}xT}-Y_{uT})2^{-j}dx\Big)^2du\Big)\\
\leq& C2^j\Big[\int_{-2^j}^{0}\Big(\mathbb{E}\Big(\int_{-2^{-j}x}^{1}\mathbf{1}_{|x|\in[0,1)}(Y_{uT+2^{-j}xT}-Y_{uT})^2du\Big)\Big)^{1/2}dx\\
&\quad+\int_0^{2^j}\Big(\mathbb{E}\Big(\int_0^{1-2^{-j}x}\mathbf{1}_{|x|\in[0,1)}(Y_{uT+2^{-j}xT}-Y_{uT})^2du\Big)\Big)^{1/2}dx\Big]^2.
\end{aligned}$$

As we took $j\geq 1$, we have $2^j>1$ in each integral, so that the first term in the last sum is 0, and

$$\begin{aligned}
&\mathbb{E}\Big(\sum_{i=1}^{n}\Big(\Delta_n^{-1}\int_{t_{i-1}}^{t_i}(Y_t-Y_{t_{i-1}})dt\Big)^2\Big)\\
\leq& C2^j\Big[\int_0^1\Big(\mathbb{E}\Big(\int_0^{1-2^{-j}x}(Y_{uT+2^{-j}xT}-Y_{uT})^2du\Big)\Big)^{1/2}dx\Big]^2\\
\leq& C2^j\Big[\int_0^1\omega_{2^{-j}xT}(Y)dx\Big]^2\\
=& C2^{j(1-2s)}T^{2s}\Big[\int_0^1 x^s(2^{-j}xT)^{-s}\omega_{2^{-j}xT}(Y)dx\Big]^2.
\end{aligned}$$

Now, $(2^{-j}xT)^{-s}\omega_{2^{-j}xT}(Y)\leq\sup_{t\in[0,T]}t^{-s}\omega_t(Y)$, which we assumed to be finite in Assumption 2.1. Say it is less than some constant K, so that

$$\mathbb{E}\Big(\sum_{i=1}^{n}\Big(\Delta_n^{-1}\int_{t_{i-1}}^{t_i}(Y_t-Y_{t_{i-1}})dt\Big)^2\Big)\leq 4C2^{j(1-2s)}T^{2s}\Big[\int_0^1 x^sK\,dx\Big]^2\leq 4C2^{j(1-2s)}T^{2s}\Big(\frac{K}{s+1}\Big)^2,$$

which goes to 0 because we assumed that $s>1/2$.

Finally, $S_n^1+S_n^2\to 0$ in probability, which means that $\Delta_n^{-1}\sum_{i=1}^{n}\overline{\Delta}_i^n(Y)\int_{(i-1)\Delta_n}^{i\Delta_n}Z_tdB_t$ converges in probability to $\int_0^T Y_tZ_tdB_t$.

2.5.2 Technical lemmas

Lemma 2.4.2. *Let* $0 < \alpha < \beta < \gamma$*. The functions* $f : x \mapsto \frac{e^{-\beta x}-1}{e^{-\alpha x}-1}$ *and* $g : x \mapsto \frac{e^{-\gamma x}-e^{-\beta x}}{e^{-\beta x}-e^{-\alpha x}}$*, defined on* $(0, +\infty)$*, are decreasing.*

To prove that lemma, let us differentiate f at $x > 0$:

$$f'(x) = \frac{-\beta e^{-\beta x}(e^{-\alpha x} - 1) + \alpha e^{-\alpha x}(e^{-\beta x} - 1)}{(e^{-\alpha x} - 1)^2}.$$

Then, $f'(x) = 0$ if and only if the numerator is zero, which is equivalent to

$$\frac{\beta}{\alpha} e^{-(\beta-\alpha)x} = f(x).$$

Assume that f is increasing over a given subinterval of $(0, +\infty)$: because $\lim_{x \to 0_+} f'(x) = \frac{\beta}{\alpha}(-\beta + \alpha) < 0$, there exists $0 < x_1 < x_2$ such that $f(x_1) \leq f(x_2)$, and $f'(x_1) = f'(x_2) = 0$. Thus, writing the previous equation for $x = x_1$ and $x = x_2$, then substracting the first equation to the second one, leads to

$$\frac{\beta}{\alpha}\left(e^{-(\beta-\alpha)x_2} - e^{-(\beta-\alpha)x_1}\right) = f(x_2) - f(x_1),$$

of which LHS is negative and RHS is positive. The function f thus cannot be increasing over any interval, which proves the result.

Then $g(x) = \frac{e^{-\gamma x}-e^{-\alpha x}}{e^{-\beta x}-e^{-\alpha x}} - 1 = \frac{e^{-(\gamma-\alpha)x}-1}{e^{-(\beta-\alpha)x}-1} - 1$. Using the first part of the lemma we get that g is decreasing too.

Lemma 2.4.3. *Under Assumptions 2.1 and 2.2 with $\alpha > 1/2$, the difference*

$$\Delta_n^{1/2} \sum_{i=1}^n \widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}} - \frac{T_2 - T_1}{e^{\vartheta(T_2-T_1)} - 1} \Delta_n^{1/2} \sum_{i=1}^n \chi_i^n,$$

with

$$\chi_i^n = \frac{\sigma_{t_{i-1}} \Delta_i^n \overline{B} \int_{t_{i-1}}^{t_i} e^{-\vartheta(T_1-t)} dB_t}{\Delta_n \overline{\sigma}_{(i-1)\Delta_n}},$$

converges to 0 *in probability.*

To prove the lemma, we show that

$$\Delta_n^{1/2} \sum_{i=1}^n \mathbb{E}\Big(\Big|\Big(\widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}} - \frac{T_2 - T_1}{e^{\vartheta(T_2-T_1)} - 1} \chi_i^n\Big)\Big|\Big) \to 0$$

in probability. Indeed, by Markov inequality, for all $\varepsilon > 0$,

$$\begin{aligned}
&\mathbb{P}\Big(\Big|\Delta_n^{1/2} \sum_{i=1}^n \widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}} - \frac{T_2 - T_1}{e^{\vartheta(T_2-T_1)} - 1} \Delta_n^{1/2} \sum_{i=1}^n \chi_i^n\Big| > \varepsilon\Big) \\
\leq & \frac{1}{\varepsilon} \mathbb{E}\Big(\Big|\Delta_n^{1/2} \sum_{i=1}^n \widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}} - \frac{T_2 - T_1}{e^{\vartheta(T_2-T_1)} - 1} \Delta_n^{1/2} \sum_{i=1}^n \chi_i^n\Big|\Big) \\
\leq & \frac{1}{\varepsilon} \Delta_n^{1/2} \sum_{i=1}^n \mathbb{E}\Big(\Big|\widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}} - \frac{T_2 - T_1}{e^{\vartheta(T_2-T_1)} - 1} \chi_i^n\Big|\Big).
\end{aligned}$$

To do so, write

$$\widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}} - \frac{T_2 - T_1}{e^{\vartheta(T_2-T_1)} - 1} \chi_i^n = \frac{(T_2 - T_1) e^{-\vartheta(T_2-T_1)}}{1 - e^{-\vartheta(T_2-T_1)}} \frac{A_i^n + B_i^n + C_i^n}{\Delta_n \overline{\sigma}_{(i-1)\Delta_n} \int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t^2 dt},$$

where

$$\begin{aligned}
A_i^n =& \int_{(i-1)\Delta_n}^{i\Delta_n} e^{-\vartheta(T_1-t)} (\sigma_t - \sigma_{(i-1)\Delta_n}) dB_t \int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t d\overline{B}_t \Delta_n \overline{\sigma}_{(i-1)\Delta_n}, \\
B_i^n =& \sigma_{(i-1)\Delta_n} \int_{(i-1)\Delta_n}^{i\Delta_n} e^{-\vartheta(T_1-t)} dB_t \Delta_n \overline{\sigma}_{(i-1)\Delta_n} \int_{(i-1)\Delta_n}^{i\Delta_n} (\overline{\sigma}_t - \overline{\sigma}_{(i-1)\Delta_n}) d\overline{B}_t, \\
C_i^n =& \sigma_{(i-1)\Delta_n} \Delta_i^n \overline{B} \int_{(i-1)\Delta_n}^{i\Delta_n} e^{-\vartheta(T_1-t)} dB_t \big(\Delta_n \overline{\sigma}^2_{(i-1)\Delta_n} - \int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t^2 dt\big).
\end{aligned}$$

By triangular inequality and using the fact that $\overline{\sigma}$ is assumed to be bounded below by $\tilde{c} > 0$, by localization,

$$\mathbb{E}\Big(\Big|\widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}} - \frac{T_2 - T_1}{e^{\vartheta(T_2-T_1)} - 1} \chi_i^n\Big|\Big) \leq \frac{1}{\Delta_n^2 \tilde{c}^3} \frac{(T_2 - T_1) e^{-\vartheta(T_2-T_1)}}{1 - e^{-\vartheta(T_2-T_1)}} \big(\mathbb{E}(|A_i^n|) + \mathbb{E}(|B_i^n|) + \mathbb{E}(|C_i^n|)\big).$$

Using localization, assume there is some $M_\Sigma > 0$ such that $\sigma_t, \overline{\sigma}_t < M_\Sigma$ for $t \in [0, T]$. Using Cauchy-Schwarz inequality,

$$\begin{aligned}
\mathbb{E}(|A_i^n|) \leq & \Delta_n M_\Sigma \sqrt{\mathbb{E}\Big(\Big(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-\vartheta(T_1-t)}(\sigma_t - \sigma_{(i-1)\Delta_n})dB_t\Big)^2\Big)\mathbb{E}\Big(\Big(\int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t d\overline{B}_t\Big)^2\Big)} \\
= & \Delta_n M_\Sigma \sqrt{\mathbb{E}\Big(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)}(\sigma_t - \sigma_{(i-1)\Delta_n})^2 dt\Big)\mathbb{E}\Big(\int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t^2 dt\Big)} \\
= & \Delta_n M_\Sigma \sqrt{\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)}\mathbb{E}\big((\sigma_t - \sigma_{(i-1)\Delta_n})^2\big) dt \mathbb{E}\Big(\int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t^2 dt\Big)} \\
\leq & (\Delta_n)^{3/2} M_\Sigma^{3/2} \sqrt{\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)}\mathbb{E}\big((\sigma_t - \sigma_{t_{i-1}})^2\big) dt}
\end{aligned}$$

by Itô isometry. Thanks to Assumption 2.2,

$$\mathbb{E}\big((\sigma_t - \sigma_{(i-1)\Delta_n})^2\big) \leq \mathbb{E}\Big(\Big(\frac{\sigma_t^2 - \sigma_{(i-1)\Delta_n}^2}{\sigma_t + \sigma_{(i-1)\Delta_n}}\Big)^2\Big) \leq \frac{1}{4\tilde{c}^2} c|t - (i-1)\Delta_n|^{2\alpha} \leq \frac{1}{4\tilde{c}^2} c\Delta_n^{2\alpha},$$

so that

$$\mathbb{E}(|A_i^n|) \leq (\Delta_n)^{2+\alpha} M_\Sigma^{3/2} \frac{\sqrt{c}}{2\tilde{c}}.$$

Using Cauchy-Schwarz, boundedness of the two volatility functions using localization then Assumption 2.2, we show in the same way that

$$\mathbb{E}(|B_i^n|) \leq (\Delta_n)^{2+\alpha} M_\Sigma^2 \frac{\sqrt{c}}{2\tilde{c}}.$$

Then, by Cauchy-Schwarz,

$$\mathbb{E}(|C_i^n|) \leq M_\Sigma \sqrt{\mathbb{E}\Big(\Big(\int_{(i-1)\Delta_n}^{i\Delta_n} (\overline{\sigma}_{(i-1)\Delta_n}^2 - \overline{\sigma}_t^2) dt\Big)^2\Big)\mathbb{E}\Big(\Big(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-\vartheta(T_1-t)} dB_t \Delta_i^n \overline{B}\Big)^2\Big)}.$$

With Jensen inequality, we have

$$\begin{aligned}
\mathbb{E}\Big(\Big(\int_{(i-1)\Delta_n}^{i\Delta_n} (\overline{\sigma}_{(i-1)\Delta_n}^2 - \overline{\sigma}_t^2) dt\Big)^2\Big) \leq & \mathbb{E}\Big(\Delta_n \int_{(i-1)\Delta_n}^{i\Delta_n} (\overline{\sigma}_{(i-1)\Delta_n}^2 - \overline{\sigma}_t^2)^2 dt\Big) \\
= & \Delta_n \int_{(i-1)\Delta_n}^{i\Delta_n} \mathbb{E}\big((\overline{\sigma}_{(i-1)\Delta_n}^2 - \overline{\sigma}_t^2)^2\big) dt \\
\leq & \Delta_n \int_{(i-1)\Delta_n}^{i\Delta_n} c\Delta_n^{2\alpha} dt \\
\leq & c\Delta_n^{2+2\alpha},
\end{aligned}$$

and

$$\mathbb{E}\Big(\Big(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-\vartheta(T_1-t)}dB_t\Delta_i^n\overline{B}\Big)^2\Big) = \mathbb{E}\Big(\Big(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-\vartheta(T_1-t)}dB_t\Big)^2\Big)\mathbb{E}\Big(\Big(\Delta_i^n\overline{B}\Big)^2\Big) \leq \Delta_n^2,$$

so that

$$\mathbb{E}(|C_i^n|) \leq M_\Sigma\sqrt{c}\Delta_n^{2+\alpha}.$$

Finally,

$$\mathbb{P}\Big(\Big|\Delta_n^{1/2}\sum_{i=1}^n \widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}} - \frac{T_2-T_1}{e^{\vartheta(T_2-T_1)}-1}\Delta_n^{1/2}\sum_{i=1}^n \chi_i^n\Big| > \varepsilon\Big) \leq \frac{K}{\varepsilon}\sum_{i=1}^n \Delta_n^{\alpha+1/2}$$

for some positive constant K. This bound converges to 0 in probability, because we assumed that $\alpha > 1/2$. This is valid for all $\varepsilon > 0$, and thus ends the proof.

Lemma 2.4.4. *Under Assumptions 2.1 and 2.2 with $\alpha > 1/2$, the sum $\Delta_n^{1/2}\sum_{i=1}^n \widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}}$ converges stably in law to a random variable defined on $(\widetilde{\Omega},\widetilde{\mathcal{F}},\widetilde{\mathbb{P}})$, which conditionally to $\mathcal{F}$ is Gaussian with variance $\widetilde{I}_{\vartheta,\sigma,\overline{\sigma}}$.*

First we establish the result for

$$\frac{T_2-T_1}{e^{\vartheta(T_2-T_1)}-1}\Delta_n^{1/2}\sum_{i=1}^n \chi_i^n,$$

with

$$\chi_i^n = \frac{\sigma_{t_{i-1}}\Delta_i^n\overline{B}\int_{t_{i-1}}^{t_i} e^{-\vartheta(T_1-t)}dB_t}{\Delta_n\overline{\sigma}_{(i-1)\Delta_n}},$$

which is done using Lemma 3.7 in [49]. To do so, we have to check conditions (3.43)–(3.46) in [49], for χ_i^n.
First, we have $\mathbb{E}(\chi_i^n|\mathcal{F}_{i-1}) = 0$, which ensures (3.43) with $A_t = 0$.
Then,

$$\begin{aligned}\mathbb{E}((\chi_i^n)^2|\mathcal{F}_{i-1}) &= \Delta_n\frac{\sigma^2_{(i-1\Delta_n}}{\Delta_n^2\overline{\sigma}^2_{(i-1)\Delta_n}}\mathbb{E}\Big(\Big(\Delta_i^n\overline{B}\Big)^2|\mathcal{F}_{i-1}\Big)\mathbb{E}\Big(\Big(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-\vartheta(T_1-t)}dB_t\Big)^2|\mathcal{F}_{i-1}\Big)\\ &= \frac{\sigma^2_{(i-1)\Delta_n}}{\overline{\sigma}^2_{(i-1)\Delta_n}}\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)}dt,\end{aligned}$$

so that

$$\sum_{i=1}^{\lfloor t/\Delta_n\rfloor}\mathbb{E}((\chi_i^n)^2|\mathcal{F}_{i-1}) \to \int_0^t \frac{e^{-2\vartheta(T_1-t)}\sigma_s^2}{\overline{\sigma}_s^2}ds$$

in probability: this is condition (3.44) with $C_t = \int_0^t \frac{e^{-2\vartheta(T_1-t)}\sigma_s^2}{\overline{\sigma}_s^2}ds$.
Then,

$$\begin{aligned}\mathbb{E}((\chi_i^n)^4|\mathcal{F}_{i-1}) &= \Delta_n^2\frac{\sigma^4_{(i-1)\Delta_n}}{\Delta_n^4\overline{\sigma}^4_{(i-1)\Delta_n}}\mathbb{E}\Big(\Big(\Delta_i^n\overline{B}\Big)^4|\mathcal{F}_{i-1}\Big)\mathbb{E}\Big(\Big(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-\vartheta(T_1-t)}dB_t\Big)^4|\mathcal{F}_{i-1}\Big)\\ &\le 9\Delta_n^2\frac{\sigma^4_{(i-1)\Delta_n}}{\overline{\sigma}^4_{(i-1)\Delta_n}}\Big(\int_{(i-1)\Delta_n}^{i\Delta_n} e^{-2\vartheta(T_1-t)}dt\Big)^2,\end{aligned}$$

as the two integrals are independent Wiener integrals. Therefore, $\sum_{i=1}^n \mathbb{E}((\chi_i^n)^4|\mathcal{F}_{i-1})$ converges to 0 in probability: this is condition (3.45).
Finally, whenever N is B, $\overline{B}$ or a bounded martingale orthogonal to $(B,\overline{B})$, we have $\mathbb{E}\big(\chi_i^n\Delta_i^n N\big|\mathcal{F}_{i-1}\big) = 0$ using independence of the processes in this product, which ensures condition (3.46).
We may thus use Lemma 3.7 in [49] to conclude, among others, that

$$\frac{T_2-T_1}{e^{\vartheta(T_2-T_1)}-1}\Delta_n^{1/2}\sum_{i=1}^n \chi_i^n$$

converges stably in law to a random variable defined on $(\widetilde{\Omega}, \widetilde{\mathcal{F}}, \widetilde{\mathbb{P}})$, which conditionally to $\mathcal{F}$ is Gaussian with variance $\widetilde{I}_{\vartheta,\sigma,\overline{\sigma}}$.

To finish, as

$$\Delta_n^{1/2} \sum_{i=1}^n \widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}} = \Delta_n^{1/2} \sum_{i=1}^n \left(\widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}} - \frac{T_2 - T_1}{e^{\vartheta(T_2 - T_1)} - 1} \sum_{i=1}^n \chi_i^n \right) + \frac{T_2 - T_1}{e^{\vartheta(T_2 - T_1)} - 1} \Delta_n^{1/2} \sum_{i=1}^n \chi_i^n,$$

the convergence of $\Delta_n^{1/2} \sum_{i=1}^n \left(\widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}} - \frac{T_2 - T_1}{e^{\vartheta(T_2 - T_1)} - 1} \sum_{i=1}^n \chi_i^n \right)$ to 0 in probability, provided by Lemma 2.4.3, and the stable convergence in law of $\frac{T_2 - T_1}{e^{\vartheta(T_2 - T_1)} - 1} \Delta_n^{1/2} \sum_{i=1}^n \chi_i^n$ that we have just proved allow to conclude that their sum converges stably in law to the same limit.

2.5.3 Some histograms from the numerical experiments

Histograms for $\vartheta = 1.4$ y^{-1}

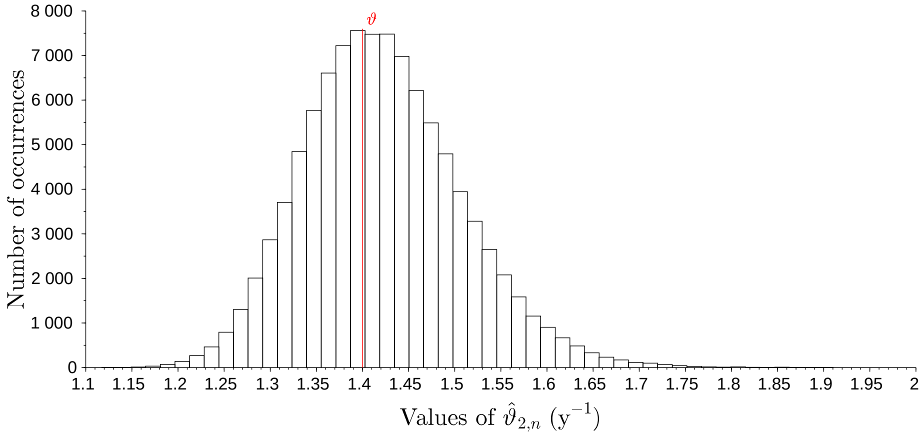

Figure 2.8 – Histogram of the 100,000 instances of $\hat{\vartheta}_{2,n}$, with 2 processes and $\vartheta = 1.4$ y^{-1}

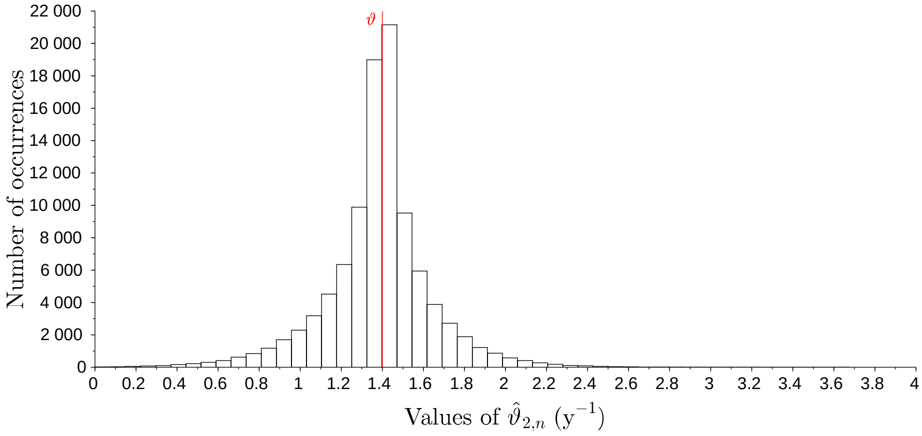

Figure 2.9 – Histogram of the 100,000 instances of $\hat{\vartheta}_{3,n}$, with 3 processes and $\vartheta = 1.4$ y^{-1}

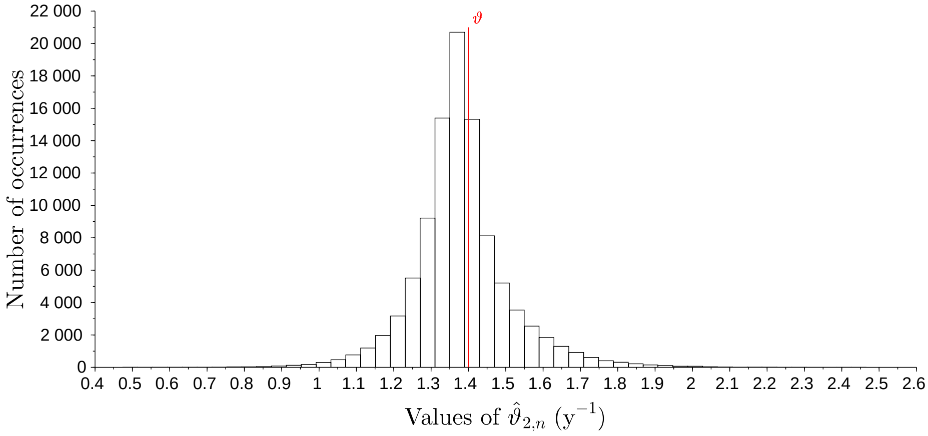

Figure 2.10 – Histogram of the 100,000 instances of $\hat{\vartheta}_{5,n}$, with 5 processes and $\vartheta = 1.4$ y^{-1}

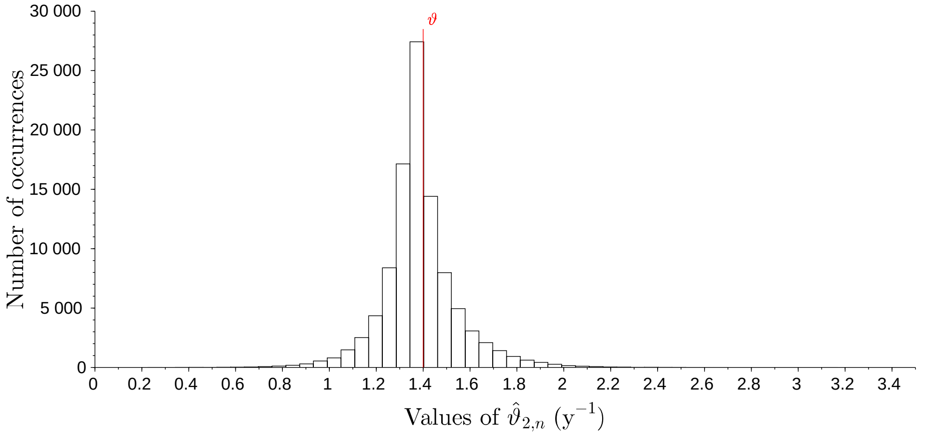

Figure 2.11 – Histogram of the 100,000 instances of $\hat{\vartheta}_{6,n}$, with 6 processes and $\vartheta = 1.4$ y^{-1}

Histograms for $\vartheta = 40$ y^{-1}

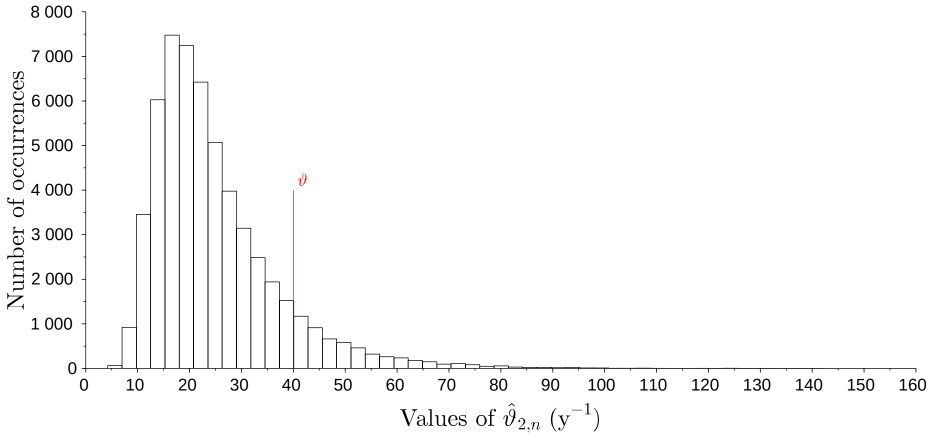

Figure 2.12 – Histogram of the 100,000 instances of $\hat{\vartheta}_{2,n}$, with 2 processes and $\vartheta = 40$ y^{-1}

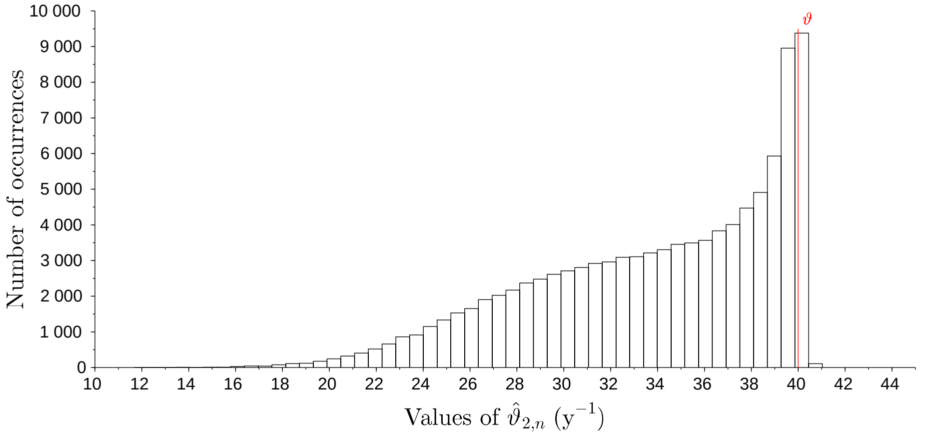

Figure 2.13 – Histogram of the 100,000 instances of $\hat{\vartheta}_{3,n}$, with 3 processes and $\vartheta = 40$ y^{-1}

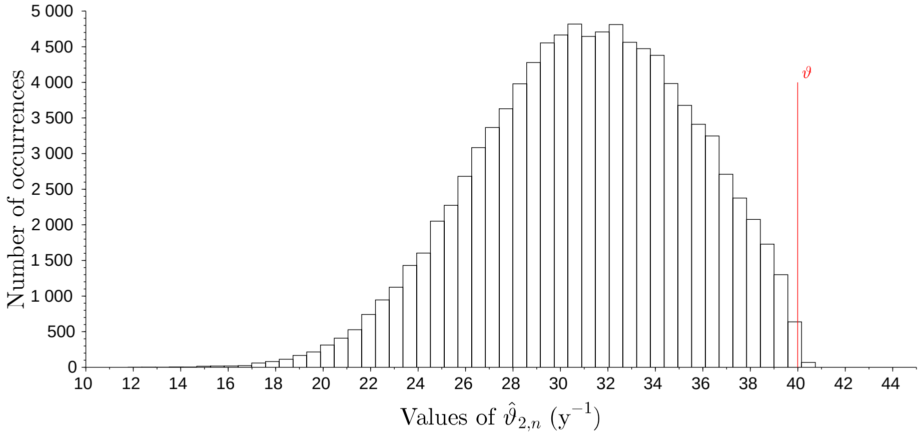

Figure 2.14 – Histogram of the 100,000 instances of $\hat{\vartheta}_{5,n}$, with 5 processes and $\vartheta = 40$ y^{-1}

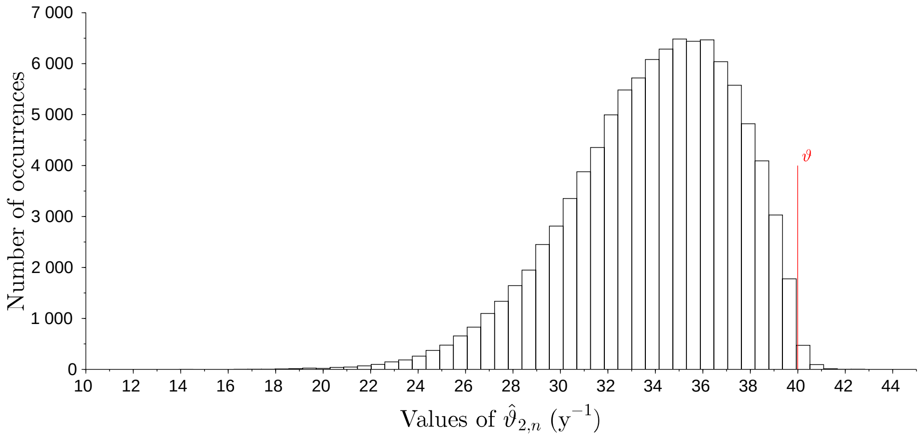

Figure 2.15 – Histogram of the 100,000 instances of $\hat{\vartheta}_{6,n}$, with 6 processes and $\vartheta = 40$ y^{-1}

Chapitre 3

Estimation in a two-factor model incorporating model errors

3.1 Introduction

3.1.1 Motivation

In this chapter, we deal with estimation procedures for noisy diffusion processes. It is quite natural to consider the possibility that a dataset does not perfectly fit the mathematical model that one applies. One motivation can come from the way data were got, as, for instance, some measurement or rounding errors may have occurred. One other reason to consider noise is that we acknowledge our model does not perfectly fit the dataset, and we therefore include model errors that fill the gap between the observables and the outputs of the mathematical model.
Estimation of processes contaminated by noise has been widely studied, a great number of applications coming from finance, where microstructure noise is a point of central interest; for instance, Zhang *et al.* [77] and then Zhang [76] have developed a "two-scale" and then a "multi-scale" approach to get a rate-optimal estimator of integrated volatility in that context. Other types of errors have retained the attention of researchers, like shrinking rounding errors in Delattre and Jacod [27]. Jacod and Protter, in the 16^{th} chapter of the book [52], get laws of large numbers and central limit theorems under a very general specification of error terms. We also refer to the bibliographical notes of that chapter for the historical evolution of the treatment of errors.

In this chapter, we use a context with shrinking noise, which stands for model errors. We consider again the context of Chapter 2, as we are modeling a multidimensional diffusion process driven by two independent Brownian motions, with the same volatility structure. We have reported that as soon as the dimension of the process is greater than two, the model is somehow degenerate: this is a common feature of Heath-Jarrow-Morton models, introduced in Heath *et al.* [40]. Jeffrey *et al.* [55] calls this phenomenon *stochastic singularity*, when there are more processes than Brownian motions. In the absence of drift processes, arbitrage would be possible as some linear combination of processes would be zero; this is not a feature of empirical data. The classic approach to avoid it is to add another source of randomness, as is done in Jeffrey *et al.* [55], Bhar and Chiarella [18] and Bhar *et al.* [19]. In the latter, estimation bearing on the prices of interest rates products is performed (in a parametric setting) with the addition of a measurement error to face stochastic singularity.

While considering noise around the diffusion process, we want to extend the results of the previous chapter, to be able to perform estimation of the parametric and nonparametric components of the volatility when data are noisy; to do so, we need the noise not to be asymptotically bigger than the process of interest, in the sense that we want it to be $O_{\mathbb{P}}(n^{-1/2})$. Such a specification will allow us to give simple extensions of the previous results, based on approximation of quadratic variation. Estimation at the rate $n^{1/2}$ will be possible, while the best rate is lower when errors are bigger and tools based on quadratic variation are usually not suitable. See Gloter and Jacod [36] for the attainable rate in a simple model with shrinking errors, and works like Jacod *et al.* [50] and the 16^{th} chapter of the book of Jacod and Protter [52] for alternative estimation procedures.
As we restrain ourselves to errors of relatively small asymptotic size, we may wonder if the model errors in the true dataset may correspond to such design. Numerical experiments shall give us some ideas about that point, while we will try to explain the gaps between the various estimators in the previous chapter when applied to electricity forward contracts prices in Section 2.3.3.

3.1.2 Setting

The basic setting is the one of the first chapter, see Section 2.1.2 : on some filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, we consider a d-dimensional Itô semimartingale $X = (X_t)_{t\geq 0}$ with components X^j for $j = 1, \dots, d$, of the form

$$X_t^j = X_0^j + \int_0^t b_s^j ds + \int_0^t e^{-\vartheta(T_j - s)}\sigma_s dB_s + \int_0^t \overline{\sigma}_s d\overline{B}_s,$$

where $X_j^0 \in \mathbb{R}$ is an initial condition, $B = (B_t)_{t\geq 0}$ and $\overline{B} = (\overline{B}_t)_{t\geq 0}$ are two independent Brownian motions, ϑ and T_j are positive numbers and $\sigma = (\sigma_t)_{t\geq 0}$, $\overline{\sigma} = (\overline{\sigma}_t)_{t\geq 0}$ and $b^j = (b_t^j)_{t\geq 0}$ are càdlàg adapted processes.

We assume that for some $T > 0$, we have

$$T \leq T_1 < \ldots < T_d$$

and that the T_i are known. The observation times are

$$0, \Delta_n, 2\Delta_n, \ldots, n\Delta_n = T,$$

and at each of the observation times, we observe X with some noise. Precisely, at time $t_i = i\Delta_n$, we have observations $Y_{t_i}^j$, $j = 1, \dots, d$, with

$$Y_{t_i^p}^j = X_{t_i}^j + \kappa_j^n \epsilon_i^j,$$

where $\kappa_j^n > 0$ are deterministic, and ϵ_i^j are iid centered random variables.
As in Chapter 2, asymptotics are taken as $n \to \infty$, and in this high-frequency framework, we are estimating ϑ and the random components $t \rightsquigarrow \sigma_t^2$ and $t \rightsquigarrow \overline{\sigma}_t^2$. In some cases, the estimators that we derived in Chapter 2 will allow us to do so with the usual rates of convergence. Depending on the structure of the error terms $\kappa_j^n \epsilon^j$, the properties of the estimators may change, in terms of asymptotic behaviour.

3.1.3 Main results

In Section 3.2.1, we give estimators of ϑ under the assumption that $n^\beta \kappa_j^n \to \iota_j$ for $j = 1, \ldots, d$ and $\beta \geq 1/2$. When $d = 2$ processes are observed, the estimator $\hat{\vartheta}_{2,n}$ of Chapter 2, which is now computed with the observations of Y instead of X, is still consistent when $\beta > 1/2$. In Theorem 3.1, we shall state that

$$\Delta_n^{-1/2}(\hat{\vartheta}_{2,n} - \vartheta) \to \mathcal{N}(0, V_\vartheta(\sigma, \overline{\sigma}))$$

stably in law when $\beta > 3/4$, where $\mathcal{N}(0, V_\vartheta(\sigma, \overline{\sigma}))$ is the same stable limit as the one of Theorem 2.1. When $1/2 < \beta < 3/4$, $\Delta_n^{1-2\beta}(\hat{\vartheta}_{2,n} - \vartheta)$ converges in probability to some random variable $M_{\vartheta,\beta}$ depending on ι_1 and ι_2, which define the limit behaviour of the error terms, and in the case $\beta = 3/4$, $\Delta_n^{-1/2}(\hat{\vartheta}_{2,n} - \vartheta)$ converges stably in law to $\mathcal{N}(0, V_\vartheta(\sigma, \overline{\sigma})) + M_{\vartheta,\beta}$.
Next we will prove that the estimators $\hat{\vartheta}_{d,n}$, for $d \geq 3$, are still consistent when $\beta > 1/2$. We shall retrieve convergence in probability, at the rate Δ_n^{-1}, to the same random variable as in the previous chapter when $\beta > 1$, while convergence in probability to another random variable depending on $\iota_1, \ldots, \iota_d$ occurs at the rate $\Delta_n^{1-2\beta}$ when $1/2 < \beta < 1$. The central case $\beta = 1$ is more difficult to treat, and we only give the result for deterministic drift and volatility processes.
Afterwards, we will introduce an estimator $\bar{\vartheta}_{3,n}$, which, at the price of the regularity assumption 2.2 with $\alpha > 1/2$ on the volatility processes and of the assumption that the volatility processes are bounded away from zero with high probability, estimates ϑ consistently when $\beta \geq 1/2$, and we will state that

$$\Delta_n^{-1/2}(\bar{\vartheta}_{3,n} - \vartheta) \to \mathcal{N}(0, V_{\vartheta,3}(\sigma, \overline{\sigma}))$$

stable in law, where $\mathcal{N}(0, V_{\vartheta,3}(\sigma, \overline{\sigma}))$ is a random variable defined on an extension of $(\Omega, \mathcal{F}, \mathbb{P})$, and which, conditionally to $\mathcal{F}$, is Gaussian, centered, with variance $V_\vartheta(\sigma, \overline{\sigma})$. Yet, that convergence occurs only for $\beta > 1/2$.

Nonparametric estimators will be extended in Section 3.2.2 in the case $d = 2$. First we will give a hint to estimate the realizations of σ_t^2 and $\overline{\sigma}_t^2$ with the optimal rates that correspond to their own regularities (and not to the lowest value of their two regularities, as did Theorem 2.2). Then we will show that the estimators $\widehat{\sigma}_{n,t}^2$ and $\widehat{\overline{\sigma}}_{n,t}^2$ of the previous chapter do not always reach the optimal rate of convergence $\Delta_n^{-\alpha/(2\alpha+1)}$, where α is the regularity of the volatility processes given by Assumption 2.2. Instead, we shall state in Theorem 3.4 that

with $\beta > 1/2$, their convergence rate is now $\Delta_n^{-(\frac{\alpha}{2\alpha+1}\wedge(2\beta-1))}$: model errors have an influence on the estimation procedure.
Yet, the estimator $\bar{\vartheta}_{3,n}$ of ϑ, which uses three processes to perform estimation, will allow us to propose new nonparametric estimators $\widehat{\sigma}^2_{3,n,t}$ and $\widehat{\overline{\sigma}}^2_{3,n,t}$ of σ_t^2 and $\overline{\sigma}_t^2$, which achieve the optimal rate as soon as $\alpha > 1/2$ and $\beta > 1/2$.

The last result is about efficient estimation; we give conditions under which the efficient estimator of the previous chapter still converges to the best possible limit, as defined when stating Theorem 2.4, at the rate $\Delta_n^{-1/2}$. In Theorem 3.6, we state that this can be done with $\alpha > 1/2$, $\beta > 3/4$ and with the volatility processes being bounded away from zero with high probability. Whether those conditions may be weakened with another proof method is an open question.

Finally, in Section 3.3, we perform numerical experiments on real data, and we use simulated datasets to compare the behaviours of the extended estimators under various specifications for model errors. This allows us to reproduce the gaps that are observed on real data between the estimators with two processes and the ones with more processes. Although we will not be able to state under which asymptotic regime we are concerning errors, we shall observe that some values of $\beta > 1/2$ are plausible.

3.2 Construction of the estimators and convergence results

3.2.1 Estimators of ϑ

In all that follows, we will work under the assumption that there are some $\beta \geq 1/2$ and some real numbers $\iota_1, \dots, \iota_d$ such that

$$n^\beta \kappa_j^n \to \iota_j$$

as $n \to \infty$. The number β will be seen as a tuning parameter defining the relative size of the noise with report to the process of interest. We will not consider the case $0 \leq \beta < 1/2$; in that context, the asymptotic size of the noise would be greater than the one of the process X. Yet estimation would still be feasible at a rate lower than the usual $\sqrt{n}$, using tools from the literature on microstructure noise. The issues that would be raised are beyond our scope. One may refer to Gloter and Jacod [36] for the best attainable rate for estimation, and to works like Jacod *et al.* [50] or Zhang [76] for estimation methods in a microstructure noise context.

The case $d = 1$

As we explained in the previous chapter, it is impossible to identify ϑ from the data in that context.

The case $d = 2$

If $\beta > 1/2$, the convergences

$$\sum_{i=1}^{n}(\Delta_i^n Y^j)^2 \to \int_0^T (e^{-2\vartheta(T_j-t)}\sigma_t^2 + \overline{\sigma}_t^2)dt, j = 1, 2 \tag{3.1}$$

and

$$\sum_{i=1}^{n}(\Delta_i^n Y^2 - \Delta_i^n Y^1)^2 \to \int_0^T (e^{-\vartheta T_2} - e^{-\vartheta T_1})^2 \sigma_t^2 dt$$

in probability remain valid. We thus expect the estimator $\hat{\vartheta}_{2,n}$ of the first chapter to remain consistent for ϑ, that is

$$\widetilde{\Psi}^n_{T_1,T_2} = \frac{\sum_{i=1}^n (\Delta_i^n Y^2 - \Delta_i^n Y^1)^2}{\sum_{i=1}^n \big((\Delta_i^n Y^2)^2 - (\Delta_i^n Y^1)^2\big)} \to \frac{(e^{-\vartheta T_2} - e^{-\vartheta T_1})^2}{e^{-2\vartheta T_2} - e^{-2\vartheta T_1}} = \psi_{T_1,T_2}(\vartheta)$$

in probability. We thus let

$$\hat{\vartheta}_{2,n} = \psi^{-1}_{T_1,T_2}\big(\widetilde{\Psi}^n_{T_1,T_2}\big)$$

if $\widetilde{\Psi}^n_{T_1,T_2} \in (-1, 0)$, and 0 else.
Yet, if $\beta = 1/2$, the convergence (3.1) becomes

$$\sum_{i=1}^{n}(\Delta_i^n Y^j)^2 \to \int_0^T (e^{-2\vartheta(T_j-t)}\sigma_t^2 + \overline{\sigma}_t^2)dt + 2\iota_j^2, j = 1, 2.$$

As we do not know the ι_j, using such sums in the estimator $\hat{\vartheta}_{2,n}$ would introduce a bias in the estimation. Only the summation

$$\sum_{i=1}^{n} \Delta_i^n Y^1 \Delta_i^n Y^2 \to \int_0^T (e^{-\vartheta(T_2+T_1-2t)}\sigma_t^2 + \overline{\sigma}_t^2)dt \tag{3.2}$$

allows to get rid of the error terms. This will not be enough to be able to estimate ϑ using tools based on quadratic variation.

The case $d \geq 3$

When $\beta > 1/2$, we use the estimator $\hat{\vartheta}_{d,n}$ introduced in the previous chapter, now computed with the observations of Y ; we shall show that

$$\widetilde{\Psi}^n_{T_{1..d}} = \sum_{j=3}^{d} \frac{\sum_{i=1}^n (\Delta_i^n Y^j - \Delta_i^n Y^{j-1})^2}{\sum_{i=1}^n (\Delta_i^n Y^2 - \Delta_i^n Y^1)^2} \to \sum_{j=3}^{d} \Big(\frac{e^{-\vartheta T_j} - e^{-\vartheta T_{j-1}}}{e^{-\vartheta T_2} - e^{-\vartheta T_1}}\Big)^2 = \psi_{T_{1..d}}(\vartheta)$$

in probability, where $\psi_{T_{1..d}}$ is a decreasing function, being the sum of $d-2$ decreasing functions (see Lemma 3.6.1). It maps $(0, +\infty)$ onto $\big(0, \sum_{j=3}^d \big(\frac{T_j - T_{j-1}}{T_2 - T_1}\big)^2\big)$. We thus choose the estimator

$$\hat{\vartheta}_{d,n} = \psi^{-1}_{T_{1..d}}\big(\widetilde{\Psi}^n_{T_{1..d}}\big)$$

whenever $\widetilde{\Psi}^n_{T_{1..d}} \in \big(0, \sum_{j=3}^d \big(\frac{T_j - T_{j-1}}{T_2 - T_1}\big)^2\big)$, and 0 otherwise.

When three processes are available, it is possible to write other summations like the one in (3.2). Observe that we have, for instance,

$$\sum_{i=1}^n \Delta_i^n Y^1(\Delta_i^n Y^3 - \Delta_i^n Y^2) \to e^{-\vartheta T_1}\big(e^{-\vartheta T_3} - e^{-\vartheta T_2}\big) \int_0^T e^{2\vartheta t}\sigma_t^2 dt$$

in probability even if $\beta = 1/2$, so that

$$\Phi^n_{T_1,T_2,T_3} = \frac{\sum_{i=1}^n \Delta_i^n Y^1(\Delta_i^n Y^2 - \Delta_i^n Y^3)}{\sum_{i=1}^n \Delta_i^n Y^2(\Delta_i^n Y^1 - \Delta_i^n Y^3)} \to \frac{e^{-\vartheta T_1}(e^{-\vartheta T_2} - e^{-\vartheta T_3})}{e^{-\vartheta T_2}(e^{-\vartheta T_1} - e^{-\vartheta T_3})} = \phi_{T_1,T_2,T_3}(\vartheta)$$

in probability, where ϕ_{T_1,T_2,T_3}, is increasing (see Lemma 3.6.1) and thus invertible. As $\vartheta \rightsquigarrow \phi_{T_1,T_2,T_3}(\vartheta)$ maps $(0,\infty)$ onto $\big(\frac{T_3-T_2}{T_3-T_1}, 1\big)$, we define the estimator $\bar{\vartheta}_{3,n}$ as

$$\bar{\vartheta}_{3,n} = \phi^{-1}_{T_1,T_2,T_3}\big(\Phi^n_{T_1,T_2,T_3}\big)$$

if $\Phi^n_{T_1,T_2,T_3} \in \big(\frac{T_3-T_2}{T_3-T_1}, 1\big)$, and $\bar{\vartheta}_{3,n} = 0$ else.

Convergence results

We enforce Assumption 2.1 of the previous chapter, and also need an assumption on the error terms:

Assumption 3.1. *Random variables ϵ_i^j are independent from each other and independent of $\mathcal{X}$, which is the filtration generated by the processes X, σ, $\overline{\sigma}$ and b. They have finite fourth-order moment, and their first- and second-order moments are*

$$\mathbb{E}(\epsilon_i^j) = 0 \text{ and } \mathbb{E}\big((\epsilon_i^j)^2\big) = 1.$$

Moreover, there exists $\iota_1, \dots, \iota_d > 0$ such that for all $j = 1, \dots, d$,

$$n^\beta \kappa_j^n \to \iota_j$$

as $n \to \infty$.

With the same notation as in the previous chapter:

Theorem 3.1. *Work under Assumptions 2.1 and 3.1, with $\beta > 1/2$. We have*

1. *if $1/2 < \beta < 3/4$,*
$$\Delta_n^{1-2\beta}(\hat{\vartheta}_{2,n} - \vartheta) \to M_{\vartheta,\beta}$$
in probability,
2. *if $3/4 < \beta$,*
$$\Delta_n^{-1/2}(\hat{\vartheta}_{2,n} - \vartheta) \to \mathcal{N}(0, V_\vartheta(\sigma, \overline{\sigma}))$$
in distribution,

3. *if* $\beta = 3/4$,
$$\Delta_n^{-1/2}(\hat{\vartheta}_{2,n} - \vartheta) \to M_{\vartheta,\beta} + \mathcal{N}(0, V_\vartheta(\sigma, \overline{\sigma}))$$
in distribution,

where
$$M_{\vartheta,\beta} = \frac{T^{1-2\beta}(e^{\vartheta T_1} + e^{\vartheta T_2})\big((\psi_{T_1,T_2}(\vartheta) - 1)\iota_2^2 + (\psi_{T_1,T_2}(\vartheta) + 1)\iota_1^2\big)}{(T_2 - T_1)(e^{-\vartheta T_2} - e^{-\vartheta T_1}) \int_0^T e^{2\vartheta t}\sigma_t^2 dt},$$
and $\mathcal{N}(0, V_\vartheta(\sigma, \overline{\sigma}))$ *is a random variable which, conditionally to* $\mathcal{F}$*, is centered normal with variance*
$$V_\vartheta(\sigma, \overline{\sigma}) = \frac{1}{(T_2 - T_1)^2}(e^{\vartheta T_2} - e^{\vartheta T_1})^2 \frac{\int_0^T e^{2\vartheta t}\sigma_t^2 \overline{\sigma}_t^2 dt}{\big(\int_0^T e^{2\vartheta t}\sigma_t^2 dt\big)^2}.$$

It appears that the estimator $\hat{\vartheta}_{2,n}$ has the same asymptotic law than in the previous chapter only when $\beta > 3/4$. Else, estimation is affected and the limit is not centered, due to the term $M_{\vartheta,\beta}$.

For the next theorem, we introduce the notation
$$\overline{b}_t^d = 2(e^{-\vartheta T_2} - e^{-\vartheta T_1}) \sum_{j=3}^d (e^{-\vartheta T_j} - e^{-\vartheta T_{j-1}})\big((e^{-\vartheta T_2} - e^{-\vartheta T_1})(b_t^j - b_t^{j-1}) - (e^{-\vartheta T_j} - e^{-\vartheta T_{j-1}})(b_t^2 - b_t^1)\big)$$
and
$$\widetilde{b}_T^d = (e^{-\vartheta T_2} - e^{-\vartheta T_1})^2 \int_0^T \sum_{j=3}^d (b_t^j - b_t^{j-1})^2 dt - \sum_{j=3}^d (e^{-\vartheta T_j} - e^{-\vartheta T_{j-1}})^2 \int_0^T (b_t^2 - b_t^1)^2 dt.$$

We also set
$$D_d = \sum_{j=3}^d (e^{-\vartheta T_j} - e^{-\vartheta T_{j-1}})\big[(e^{-\vartheta T_j} - e^{-\vartheta T_{j-1}})(T_2 e^{-\vartheta T_2} - T_1 e^{-\vartheta T_1}) - (e^{-\vartheta T_2} - e^{-\vartheta T_1})(T_j e^{-\vartheta T_j} - T_{j-1} e^{-\vartheta T_{j-1}})\big].$$

Theorem 3.2. *Work under Assumptions 2.1 and 3.1 with* $\beta > 1/2$*. We have*

1. *if* $1/2 < \beta < 1$,
$$\Delta_n^{1-2\beta}(\hat{\vartheta}_{d,n} - \vartheta) \to \frac{T^{1-2\beta}(e^{-\vartheta T_2} - e^{-\vartheta T_1}) \sum_{j=3}^d \Big((\iota_{j-1}^2 + \iota_j^2) - \frac{\mathfrak{c}_{j-1,j}(\vartheta)^2}{\mathfrak{c}_{1,2}(\vartheta)^2}(\iota_1^2 + \iota_2^2)\Big)}{D_d \int_0^T e^{2\vartheta t}\sigma_t^2 dt}$$
in probability.
2. *if* $1 < \beta$,
$$\Delta_n^{-1}(\hat{\vartheta}_{d,n} - \vartheta) \to \frac{\widetilde{b}_T^d + \int_0^T \overline{b}_t^d e^{\vartheta t}\sigma_t dB_t}{2D_d(e^{-\vartheta T_2} - e^{-\vartheta T_1}) \int_0^T e^{2\vartheta t}\sigma_t^2 dt}$$
in probability.

Remark 3.1. *We did not treat the case $\beta = 1$ in Theorem 3.2, as there would be one more non-negligible term in the proof, and it would be hard to work out using our proof method, based on Lemma 3.7 in [49] and its more general version, Theorem 7.28 in [54]. Yet, if the drift and volatility processes are deterministic, we may use a more standard CLT for dependent variables like the one in [28], and we state the result without a proof, if $\beta = 1$, needing moreover that $\alpha > 1/2$. Then,*

$$\Delta_n^{-1}(\hat{\vartheta}_{d,n} - \vartheta) \to \mathcal{N}(M_1, V_1)$$

in distribution, where

$$\begin{aligned} M_1 =& \frac{T^{1-2\beta}(e^{-\vartheta T_2} - e^{-\vartheta T_1}) \sum_{j=3}^{d} \Big((\iota_{j-1}^2 + \iota_j^2) - \frac{\mathfrak{e}_{j-1,j}(\vartheta)^2}{\mathfrak{e}_{1,2}(\vartheta)^2}(\iota_1^2 + \iota_2^2) \Big)}{D_d \int_0^T e^{2\vartheta t} \sigma_t^2 dt} \\ &+ \frac{\widetilde{b}_T^d}{2D_d(e^{-\vartheta T_2} - e^{-\vartheta T_1}) \int_0^T e^{2\vartheta t}\sigma_t^2 dt} \end{aligned}$$

and with

$$\begin{aligned} V_1 =& \frac{\int_0^T \big(\overline{b}_t^d\big)^2 e^{2\vartheta t}\sigma_t^2 dt}{4D_d^2(e^{-\vartheta T_2} - e^{-\vartheta T_1})^2 \big(\int_0^T e^{2\vartheta t}\sigma_t^2 dt \big)^2} \\ &+ \frac{(e^{-\vartheta T_2} - e^{-\vartheta T_1})^2}{2D_d^2 \int_0^T e^{2\vartheta t}\sigma_t^2 dt} \Big[\sum_{j=3}^{d} \frac{\mathfrak{e}_{j-1,j}(\vartheta)^4}{\mathfrak{e}_{1,2}(\vartheta)^2} \iota_1^2 \\ &\qquad + \Big(\mathfrak{e}_{2,3}(\vartheta)^2 + \sum_{j=3}^{d} \frac{\mathfrak{e}_{j-1,j}(\vartheta)^4}{\mathfrak{e}_{1,2}(\vartheta)^2} + 2\frac{\mathfrak{e}_{2,3}(\vartheta)}{\mathfrak{e}_{1,2}(\vartheta)} \sum_{j=3}^{d} \mathfrak{e}_{j-1,j}(\vartheta)^2 \Big) \iota_2^2 \\ &\qquad + \sum_{j=3}^{d-1} \mathfrak{e}_{j-1,j+1}(\vartheta)\iota_j^2 + \mathfrak{e}_{d-1,d}(\vartheta)\iota_d^2 \Big]. \end{aligned}$$

Finally, we state the result relative to the estimator $\bar{\vartheta}_{3,n}$:

Theorem 3.3. *Work under Assumptions 2.1, 2.2 and 3.1, with $\alpha > 1/2$. Assume that $\mathbb{P}((\sigma, \overline{\sigma}) \in \Sigma(c, \tilde{c})) = 1$.*
Then $\bar{\vartheta}_{3,n}$ is a consistent estimator for ϑ if $\beta \geq 1/2$, and if $\beta > 1/2$, we have

$$\Delta_n^{-1/2}(\bar{\vartheta}_{3,n} - \vartheta) \to \mathcal{N}(0, V_{\vartheta,3}(\sigma, \overline{\sigma}))$$

in distribution as $n \to \infty$, where $\mathcal{N}(0, V_{\vartheta,3}(\sigma, \overline{\sigma}))$ is a random variable which, conditionally to $\mathcal{F}$, is centered normal with variance

$$V_{\vartheta,3}(\sigma, \overline{\sigma}) = \frac{e^{2\vartheta(T_1+T_2+T_3)}(e^{-\vartheta T_1} - e^{-\vartheta T_2})^2(e^{-\vartheta T_2} - e^{-\vartheta T_3})^2(e^{-\vartheta T_3} - e^{-\vartheta T_1})^2 \int_0^T e^{2\vartheta t}\sigma_t^2\overline{\sigma}_t^2 dt}{\Big(\big(T_1(e^{-\vartheta T_2} - e^{-\vartheta T_3}) + T_2(e^{-\vartheta T_3} - e^{-\vartheta T_1}) + T_3(e^{-\vartheta T_1} - e^{-\vartheta T_2})\big) \int_0^T e^{2\vartheta t}\sigma_t^2 dt\Big)^2}.$$

Remark 3.2. *1. A careful examination of the proof of this theorem allows us to see that the estimator is consistent even for $\beta > 1/4$, but we dropped out the case $\beta < 1/2$ in our study.*

2. In the theorem, the limit law is stated only for $\beta > 1/2$. Indeed, the conditional variance would be different when $\beta = 1/2$, and other tools would be required to establish the limit law.

3.2.2 Nonparametric estimation of the volatility processes

Best-rate pointwise estimation with no model errors, $d = 2$

In this section, we temporarily remove model errors, assuming that we observe the process X directly, and we extend the results of Chapter 2, concerning nonparametric estimation. Indeed, it was stated in Theorem 2.2 that under Assumptions 2.1 and 2.2, the sequence

$$\Delta_n^{-\alpha/(2\alpha+1)} \sup_{t\in[h_n,T]} \Big[\big|\widehat{\sigma}^2_{n,t} - \sigma^2_t\big| + \big|\widehat{\overline{\sigma}}^2_{n,t} - \overline{\sigma}^2_t\big|\Big]$$

was tight. The normalizing rate depends on α which, regarding Assumption 2.2, is the worst value of the two regularities of σ and $\overline{\sigma}$, in the following sense. Suppose we work under the following assumption:

Assumption 3.2. *There exists a constant $c > 0$ and $\alpha \geq 1/2$, $\bar{\alpha} \geq 1/2$ such that for every $t, s \in [0, T]$, we have*

$$\mathbb{E}\big[|\sigma^2_t - \sigma^2_s|^2\big] \leq c|t-s|^{2\alpha} \text{ and } \mathbb{E}\big[|\overline{\sigma}^2_t - \overline{\sigma}^2_s|^2\big] \leq c|t-s|^{2\bar{\alpha}}.$$

Then Theorem 2.2 gives tightness of the sequence

$$\Delta_n^{-(\alpha\wedge\bar{\alpha})/(2(\alpha\wedge\bar{\alpha})+1)} \sup_{t\in[h_n,T]} \Big[\big|\widehat{\sigma}^2_{n,t} - \sigma^2_t\big| + \big|\widehat{\overline{\sigma}}^2_{n,t} - \overline{\sigma}^2_t\big|\Big],$$

because when we perform the summation

$$\sum_{t-h_n\leq(i-1)/n<t} \begin{pmatrix} (\Delta_i^n X^1)^2 \\ (\Delta_i^n X^2)^2 \end{pmatrix} \to \begin{pmatrix} e^{-2\vartheta(T_1-t)}\sigma^2_t + \overline{\sigma}^2_t \\ e^{-2\vartheta(T_2-t)}\sigma^2_t + \overline{\sigma}^2_t \end{pmatrix}$$

before inverting the 2×2 linear system in the definition 2.3 of the estimators $\widehat{\sigma}^2_n$ and $\widehat{\overline{\sigma}}^2_n$, we have to choose a common bandwidth h_n which accomodates the two processes to be estimated pointwise.

Our basic idea is to perform summation after inverting the linear system, defining a new estimator, that is

$$\begin{pmatrix} \breve{\sigma}^2_{n,t} \\ \breve{\overline{\sigma}}^2_{n,t} \end{pmatrix} = \begin{pmatrix} h_n^{-1}\sum_{t-h_n\leq(i-1)/n<t} \mathcal{M}^1_i \\ \bar{h}_n^{-1}\sum_{t-\bar{h}_n\leq(i-1)/n<t} \mathcal{M}^2_i \end{pmatrix},$$

where

$$\begin{pmatrix} \mathcal{M}_i^1 \\ \mathcal{M}_i^2 \end{pmatrix} = \mathcal{M}(\max\{\hat{\vartheta}_{2,n}, \varpi_n\})_t \begin{pmatrix} (\Delta_i^n X^1)^2 \\ (\Delta_i^n X^2)^2 \end{pmatrix}.$$

As this new construction allows us to choose different bandwidths h_n and $\bar{h}_n$ to estimate σ_t^2 and $\overline{\sigma}_t^2$, we expect to be able to estimate each of the two values with their best possible rate of convergence.

Extensions of nonparametric results with model errors, $d = 2$

From now, we have model errors in the observations again. Our aim is to characterize the behaviour of the estimators $\widehat{\sigma}_n^2$ and $\widehat{\overline{\sigma}}_n^2$ (or $\breve{\sigma}_n^2$ and $\breve{\overline{\sigma}}_n^2$, that we have just introduced) when the processes X^1 and X^2 are observed with errors. Take the estimator of Chapter 2 computed with the observations of Y :

$$\begin{pmatrix} \widehat{\sigma}_{n,t}^2 \\ \widehat{\overline{\sigma}}_{n,t}^2 \end{pmatrix} = h_n^{-1} \mathcal{M}(\max\{\hat{\vartheta}_{2,n}, \varpi_n\})_t \sum_{t-h_n \le (i-1)/n < t} \begin{pmatrix} (\Delta_i^n Y^1)^2 \\ (\Delta_i^n Y^2)^2 \end{pmatrix}.$$

When the data are too noisy (even if $\beta > 1/2$), estimation at the best rate is not possible, as will be stated in Theorem 3.4. For this estimator too, it is possible to perform summation after inverting the linear system in order to use different bandwidths for σ and $\overline{\sigma}$. When noise is light, the optimal rate will be attained, whereas this hint will not cancel the effect of the errors if they are important.

Remark 3.3. *If more than three processes are available, one could think of plugging the estimator $\bar{\vartheta}_{3,n}$ in $\widehat{\sigma}_{n,t}^2$ and $\widehat{\overline{\sigma}}_{n,t}^2$, instead of $\hat{\vartheta}_{2,n}$. However, this would not lead to a better convergence rate, see the proof in Section 3.4.5 for more details.*

Rate-optimal nonparametric estimation with model errors, $d = 3$

As stated in Theorem 3.4, the estimators $\widehat{\sigma}_{n,t}^2$ and $\widehat{\overline{\sigma}}_{n,t}^2$ that we have just described do not reach the optimal rate of convergence when β is too low. We propose the estimator

$$\begin{pmatrix} \widetilde{\sigma}_{3,n,t}^2 \\ \widetilde{\overline{\sigma}}_{3,n,t}^2 \end{pmatrix} = h_n^{-1} \widetilde{\mathcal{M}}(\max\{\bar{\vartheta}_{3,n}, \varpi_n\})_t \sum_{t-h_n \le (i-1)/n < t} \begin{pmatrix} \Delta_i^n Y^1 \Delta_i^n Y^2 \\ \Delta_i^n Y^1 \Delta_i^n Y^3 \end{pmatrix},$$

with

$$\widetilde{\mathcal{M}}(\vartheta)_t = \frac{1}{e^{-\vartheta(T_1+T_2-2t)} - e^{-\vartheta(T_1+T_3-2t)}} \begin{pmatrix} 1 & -1 \\ -e^{-\vartheta(T_1+T_3-2t)} & e^{-\vartheta(T_1+T_2-2t)} \end{pmatrix}.$$

We shall see that it reaches the optimal rate of convergence, but it can only be implemented if 3 processes are available.

Convergence results

Proposition 3.1. *Work under Assumptions 2.1 and 3.2. Let h_n and $\bar{h}_n$ be specified by*

$$h_n = \Delta_n^{1/(2\alpha+1)} \text{ and } \bar{h}_n = \Delta_n^{1/(2\bar{\alpha}+1)},$$

and let ϖ_n be any sequence of positive numbers that decreases to 0.
Then the sequences

$$\Delta_n^{-\alpha/(2\alpha+1)} \sup_{t\in[h_n,T]} \left|\breve{\sigma}^2_{n,t} - \sigma^2_t\right| \text{ and } \Delta_n^{-\bar{\alpha}/(2\bar{\alpha}+1)} \sup_{t\in[h_n,T]} \left|\breve{\overline{\sigma}}^2_{n,t} - \overline{\sigma}^2_t\right|$$

are tight. This implies that

$$\Delta_n^{-\alpha/(2\alpha+1)}\left|\breve{\sigma}^2_{n,t} - \sigma^2_t\right| \text{ and } \Delta_n^{-\bar{\alpha}/(2\bar{\alpha}+1)}\left|\breve{\overline{\sigma}}^2_{n,t} - \overline{\sigma}^2_t\right|$$

are tight, uniformly for $t \in \mathcal{D}$, where $\mathcal{D}$ is any compact included in $(0,T]$.

The proof of this result is exactly the same as the proof of Theorem 2.2 in Section 2.4.3. Indeed, the expressions of the two estimators are exactly the same as in Chapter 2, except that the bandwidth of each estimator can be chosen without interfering with the other one. We therefore may lead the same proof and adopt the bandwidths $h_n = \Delta_n^{1/(2\alpha+1)}$ and $\bar{h}_n = \Delta_n^{1/(2\bar{\alpha}+1)}$ while respectively estimating σ^2_t and $\overline{\sigma}^2_t$.

Theorem 3.4. *Work under Assumptions 2.1, 2.2 and 3.1, with $\beta > 1/2$. Let h_n be specified by*

$$h_n = \Delta_n^{1/(2\alpha+1)},$$

and let ϖ_n be any sequence of positive numbers that decreases to 0.
Then the sequence

$$\Delta_n^{-\left(\frac{\alpha}{2\alpha+1}\wedge(2\beta-1)\right)} \sup_{t\in[h_n,T]} \left[\left|\widehat{\sigma}^2_{n,t} - \sigma^2_t\right| + \left|\widehat{\overline{\sigma}}^2_{n,t} - \overline{\sigma}^2_t\right|\right]$$

is tight.

Theorem 3.5. *Work under Assumptions 2.1, 2.2 and 3.1 with $\alpha > 1/2$ and $\beta > 1/2$. Assume that $\mathbb{P}((\sigma,\overline{\sigma}) \in \Sigma(c,\tilde{c})) = 1$. Let h_n be specified by*

$$h_n = \Delta_n^{1/(2\alpha+1)},$$

and let ϖ_n be any sequence of positive numbers that decreases to 0.
Then the sequence

$$\Delta_n^{-\alpha/(2\alpha+1)} \sup_{t\in[h_n,T]} \left[\left|\widetilde{\sigma}^2_{3,n,t} - \sigma^2_t\right| + \left|\widetilde{\overline{\sigma}}^2_{3,n,t} - \overline{\sigma}^2_t\right|\right]$$

is tight.

Theorems 3.4 and 3.5 can accomodate Assumption 3.2 if we invert the linear system before performing summation. For simplicity, we yet state the theorems under Assumption 2.2 only. Notice also, by looking at the proof, that having $\alpha > 1/2$, $\beta > 1/2$ and $\mathbb{P}((\sigma,\overline{\sigma}) \in \Sigma(c,\tilde{c})) = 1$ in Theorem 3.5 is only needed because we plug $\bar{\vartheta}_{3,n}$ into those nonparametric estimators. If we managed to get rid of those assumptions in Theorem 3.3 (see Remark 3.2), they could be removed from Theorem 3.5.

3.2.3 Efficient estimation of ϑ in presence of model errors, when $d = 2$

We are now concerned with Theorem 2.4 of the previous chapter; we would like to get sufficient conditions so that efficient estimation is still possible, that is, so that there exists some $\Delta_n^{-1/2}$-consistent estimator reaching the lower bound

$$V_\vartheta^{\mathrm{opt}}(\sigma,\overline{\sigma}) = \frac{1}{(T_2-T_1)^2}(e^{\vartheta T_2}-e^{\vartheta T_1})^2\Big(\int_0^T \frac{e^{2\vartheta t}\sigma_t^2}{\overline{\sigma}_t^2}dt\Big)^{-1}$$

given in Theorem 2.3. We state a result giving conditions under which this is feasible. For technical reasons, we replace the estimator $\hat{\vartheta}_{2,n}$ by $\Delta_n^{1/2}\lfloor \Delta_n^{-1/2}\hat{\vartheta}_{2,n}\rfloor$ and $\widehat{\overline{\sigma}}_{n,t}^2$ by $\max\big(\widehat{\overline{\sigma}}_{n,t}^2, \tilde{c}^2\big)$, where $\tilde{c}$ is the lower bound associated to $\Sigma(c,\tilde{c})$. We still write $\hat{\vartheta}_{2,n}$ and $\widehat{\overline{\sigma}}_{n,t}^2$ for simplicity.

Theorem 3.6. *Work under Assumptions 2.1, 2.2 and 3.1 with $\alpha > 1/2$, $\beta > 3/4$, and $\mathbb{P}\big((\sigma,\overline{\sigma}) \in \Sigma(c,\tilde{c})\big) = 1$. The estimator $\tilde{\vartheta}_{2,n}$ defined by*

$$\tilde{\vartheta}_{2,n} = \hat{\vartheta}_{2,n} + \frac{\sum_{i\in\mathcal{I}_n}\widetilde{\ell}\big(\hat{\vartheta}_{2,n},\widehat{\overline{\sigma}}_n^2\big)^i}{\sum_{i\in\mathcal{I}_n}\big(\widetilde{\ell}\big(\hat{\vartheta}_{2,n},\widehat{\overline{\sigma}}_n^2\big)^i\big)^2}$$

with $\mathcal{I}_n = \{i = 1,\dots,n | h_n \le t_{i-1} < T\}$ and

$$\widetilde{\ell}\big(\hat{\vartheta}_{2,n},\widehat{\overline{\sigma}}_n^2\big)^i = \frac{(\Delta_i^n Y^2 - \Delta_i^n Y^1)(\Delta_i^n Y^2 - e^{-\hat{\vartheta}_{2,n}(T_2-T_1)}\Delta_i^n Y^1)e^{-\hat{\vartheta}_{2,n}(T_2-T_1)}(T_2-T_1)}{(1-e^{-\hat{\vartheta}_{2,n}(T_2-T_1)})^3\Delta_n\widehat{\overline{\sigma}}_{n,t_{i-1}}^2}$$

satisfies

$$\Delta_n^{-1/2}\big(\tilde{\vartheta}_{2,n}-\vartheta\big) \to \mathcal{N}\big(0, V_\vartheta^{\mathrm{opt}}(\sigma,\overline{\sigma})\big)$$

in distribution, where $\mathcal{N}\big(0,V_\vartheta^{\mathrm{opt}}(\sigma,\overline{\sigma})\big)$ is, conditionally to $\mathcal{F}$, centered Gaussian with (conditional) variance $V_\vartheta^{\mathrm{opt}}(\sigma,\overline{\sigma})$.

Remark 3.4. *When $1/2 < \beta < 3/4$, the estimator $\hat{\vartheta}_{2,n}$ is not $\Delta_n^{-1/2}$-consistent, and we therefore do not hope to perform efficient estimation. Whether it is possible or not when $\beta = 3/4$ is still an open problem, as the tools used in the proof currently do not accomodate this limit case.*

3.3 Numerical implementation

We now perform some numerical tests to show how the estimators presented in this chapter behave on sets of simulated and real data. First we work on real data, then we will use simulated datasets to assess the effect of model errors.

3.3.1 Results on real data

We are working again with the dataset presented in Section 2.3.3 of the previous chapter. In that section, we applied all the estimators of ϑ that we got. Now we have introduced $\bar{\vartheta}_{3,n}$, and we may apply it on the 142 periods of three months that are present in the dataset. We present the number of periods on which the estimator converged, together with the average and the standard deviation of the converging instances, in Table 3.1. For comparison, we put again the values of the estimators $\hat{\vartheta}_{2,n}, \ldots, \hat{\vartheta}_{6,n}$ and $\tilde{\vartheta}_{2,n}$, which have already been given in Chapter 2 as $\hat{\vartheta}_{2,n}, \ldots, \hat{\vartheta}_{6,n}$ and $\tilde{\vartheta}_{2,n}$ (recall that they are the same ones, respectively in a context with and without model errors).

Estimator	Per. with convergence/ Number of per.	Average	Standard deviation
$\hat{\vartheta}_{2,n}$	49/141	26.065	11.788
$\tilde{\vartheta}_{2,n}$	49/141	26.081	11.779
$\bar{\vartheta}_{3,n}$	65/142	17.559	19.349
$\hat{\vartheta}_{3,n}$	100/142	4.3707	3.5329
$\hat{\vartheta}_{4,n}$	111/143	3.1333	2.6758
$\hat{\vartheta}_{5,n}$	111/139	2.0936	2.4969
$\hat{\vartheta}_{6,n}$	105/125	3.3881	2.8221

Table 3.1 – Estimators of ϑ on real data in France (unit: y)

The estimator $\bar{\vartheta}_{3,n}$ converges on 65 periods out of 142, which is better than $\hat{\vartheta}_{2,n}$ but worse than $\hat{\vartheta}_{3,n}, \ldots, \hat{\vartheta}_{6,n}$. The instances that converged have an average that is between the one of $\hat{\vartheta}_{2,n}$ and the ones of $\hat{\vartheta}_{3,n}, \ldots, \hat{\vartheta}_{6,n}$, and a standard deviation that is high, compared to all the other estimators. Moreover, we recall that its rate of convergence is $\Delta_n^{-1/2}$, like $\hat{\vartheta}_{2,n}$. Concerning the other estimators, depending on model errors, their convergence rates may be better (up to Δ_n^{-1}) or worse ($\Delta_n^{1-2\beta}$).
All this analysis is valid only when $\beta > 1/2$, so that we may get the rates of convergence from the theorems of this chapter. We may wonder if this is the case, and we would like to understand how such differences between the estimators may occur while the number of processes that are used changes.

3.3.2 Experiment on model errors using simulated data

In this part, inspired from the study on real data, we evaluate the impact of error terms on the estimation of ϑ thanks to simulation. More precisely, we consider an additive Gaussian noise ϵ with mean 0 and variance 1. Because the configurations differ in the number of observations, we propose to add a noise of which standard deviation is proportional to some power of n. It will be a function of parameters (β, χ) : for all $j = 1, \ldots, d$ and $i = 1, \ldots, n$, the standard deviation of $\kappa_j^n \epsilon_i^j$ is equal to $\kappa_j^n = \frac{10^{-\chi}}{n^\beta}$. To simulate data, we set $\sigma = 2.1865\ y^{-1/2}$ and $\overline{\sigma} = 0.23550\ y^{-1/2}$, which are the averages of the values that we estimated on the 30 last days in the experiment on real data of Section 2.3.3. Moreover, we used two different values for ϑ, 2.0936 y^{-1} and 26.065 y^{-1} (coming from estimation on real observations, see Table 3.1) and 2 values of the exponent χ (1 and 2).

Figures 3.1, 3.2, 3.3 and 3.4 show the estimators of ϑ in the four cases that are defined by the two values of ϑ and the two values of χ. In each of those four cases, we have performed 100,000 simulations for various values of β. Each graph represents the estimation results (average and empirical confidence interval) of the instances of $\hat{\vartheta}_{2,n}$, $\hat{\vartheta}_{3,n}$ and $\hat{\vartheta}_{6,n}$ that converged, together with the ones of $\bar{\vartheta}_{3,n}$, with respect to different values of β. In a given configuration, it may happen that a great percentage of the estimators do not converge (as they all rely on inverting some function, which is not always possible). Most of the time, such a situation is met for $\chi = 1$ and for low values of β. We do not plot the points corresponding to cases in which less than 15% of the instances of the considered estimator converged.

When the noise variance is small ($\chi = 2$) and for the smallest of the two values of ϑ, we see in Figure 3.1 that the averages of the four estimators are close to the true value, plotted in solid black line, when the parameter β is high. As it decreases, we see that all the estimators($\hat{\vartheta}_{2,n}$ (plotted in green), $\bar{\vartheta}_{3,n}$ (in thin black), $\hat{\vartheta}_{3,n}$ (red) and $\hat{\vartheta}_{6,n}$ (blue)) take a wider range of values, which is, yet, still centered at the true value.
For the highest value of ϑ, we see in Figure 3.2 that while the values of $\hat{\vartheta}_{2,n}$ and $\bar{\vartheta}_{3,n}$ are not very sensitive to the increasing of error, the ones of $\hat{\vartheta}_{3,n}$ and $\hat{\vartheta}_{6,n}$ decrease: it appears that these two estimators tend to be biased when β is low.

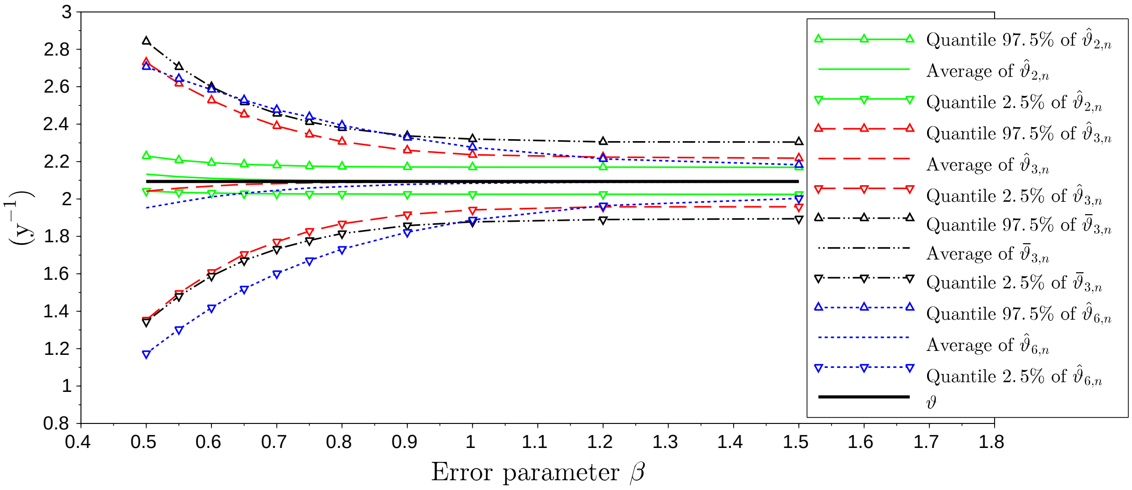

Figure 3.1 – Average and quantiles of estimators on simulated data, with $\vartheta = 2.0936$ y^{-1} and $\chi = 2$

When the noise variance is greater ($\chi = 1$), we observe different behaviours: Figure 3.3 shows that the estimator $\hat{\vartheta}_{2,n}$ takes quickly increasing values as β decreases, while the ones of $\bar{\vartheta}_{3,n}$ increase a bit, together with the length of the quantile interval. The lengths of the

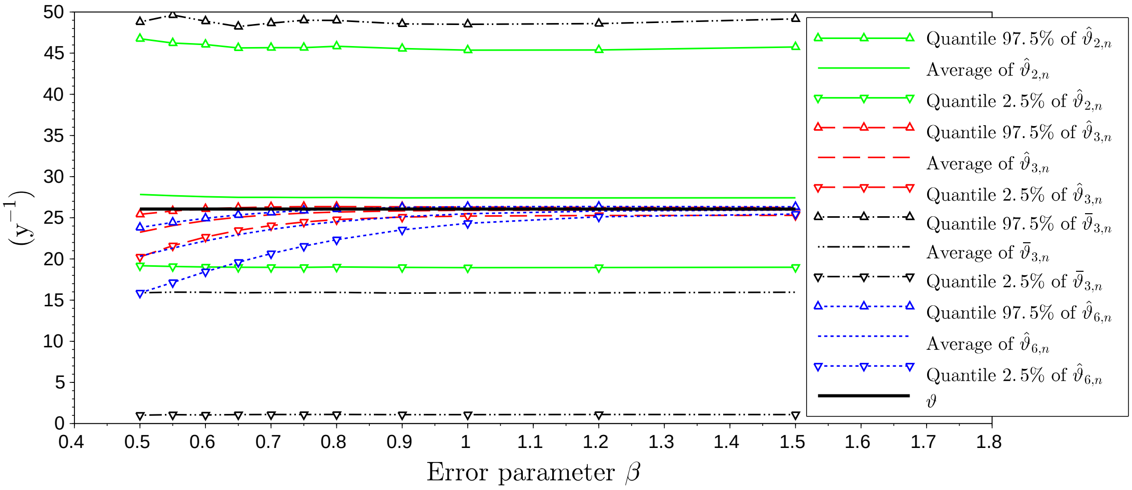

Figure 3.2 – Average and quantiles of estimators on simulated data, with $\vartheta = 26.065$ y^{-1} and $\chi = 2$

quantile intervals of $\hat{\vartheta}_{3,n}$ and $\hat{\vartheta}_{6,n}$ increase too, but these two estimator exhibit an incrasing bias downwards as β becomes lower.

In Figure 3.4 where ϑ has the highest of the two values of the experiment, we see that $\bar{\vartheta}_{3,n}$ is almost insensitive to a change of value of β, while the average and quantiles of $\hat{\vartheta}_{2,n}$ increase as β decreases. In the same time, we see an important bias downwards for the estimators $\hat{\vartheta}_{3,n}$ and $\hat{\vartheta}_{6,n}$, even if, for each of them, the difference between the two quantiles seems to be lower for low values of β.

Having performed those simulations, we may look again at the results on real data in Table 3.1. We are almost able to reproduce the gaps between the estimators on real data using simulation; in particular, Figure 3.4 shows that with a high ϑ, such differences may occur. With a low β as in Figure 3.3, the gaps do not have the same amplitude as in real data, but they could undoubtedly be reproduced with values of β less than $1/2$, or with a specification of model errors $\kappa_j^n \epsilon_i^j$, with $\kappa_j^n = \frac{k}{n^\beta}$, $k < 10^{-1}$. It is therefore uneasy to say which of the two values most likely corresponds to the real data.

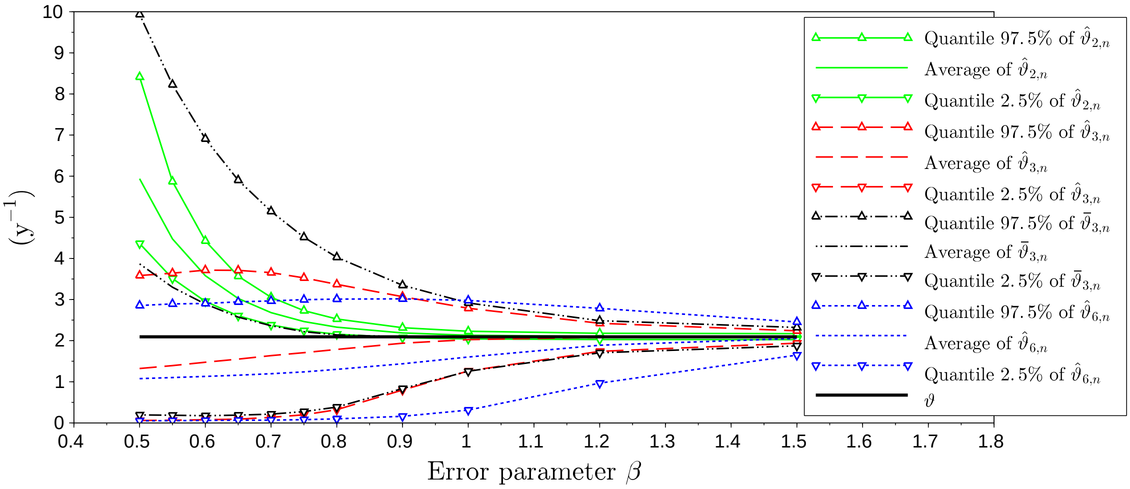

Figure 3.3 – Average and quantiles of estimators on simulated data, with $\vartheta = 2.0936$ y^{-1} and $\chi = 1$

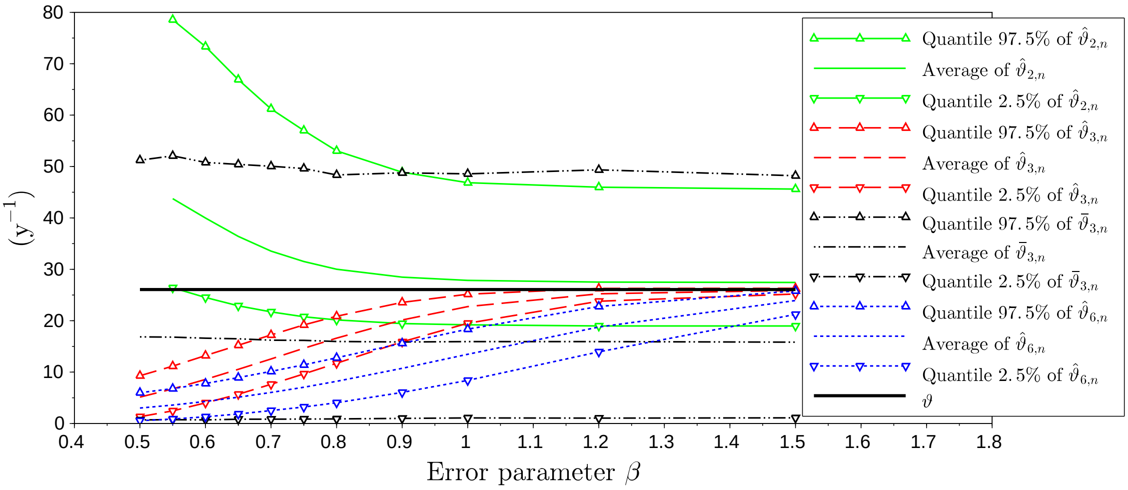

Figure 3.4 – Average and quantiles of estimators on simulated data, with $\vartheta = 26.065$ y^{-1} and $\chi = 1$

3.3.3 Impact of model errors on nonparametric estimation

In that last part, we look at the way model errors affect the nonparametric estimators. Recall that in this chapter, we proved Theorems 3.4 and 3.5 which assert that the estimators $\widetilde{\sigma}^2_{3,n,t}$ and $\widetilde{\overline{\sigma}}^2_{3,n,t}$, using $d = 3$ processes, reach the optimal rate of convergence while estimating σ^2_t and $\overline{\sigma}^2_t$. On the contrary, $\widehat{\sigma}^2_{n,t}$ and $\widehat{\overline{\sigma}}^2_{n,t}$, using $d = 2$ processes, may converge at a lower rate because of model errors: the rate of convergence will be non-optimal as soon as $\frac{\alpha}{2\alpha+1} > 2\beta - 1$.

We perform again the experiment of Section 2.3.2 : we use the causal kernel $K(x) = \mathbf{1}_{(0,1](x)}$, the bandwidth h_n is taken equal to 14 days, and the bound ϖ_n is $3.65 \cdot 10^{-2}$ y^{-1}. We simulate three processes with $T = T_1 = 150$, $T_2 = 181$ and $T_3 = 212$, with $n = 100$ observation times and $\vartheta = 10$ y^{-1}. We take $b^j_t = 3.65 \cdot 10^{-1}(\log(30) - X^j_t)$ for the drift process, and $\sigma_t = 0.37\Sigma^d_t, \overline{\sigma}_t = 0.15\Sigma^d_t$, where $\Sigma^d_t = \sqrt{\frac{1}{2}X^1_t + \frac{1}{2}X^2_t}$. We simulate Gaussian model errors $\epsilon^j_i \sim \mathcal{N}(0, 1)$, and we take $\kappa^n_j = \frac{10^{-1}}{n^\beta}$, for various values of β.
As we did before, we perform simulation and estimation 10,000 times, and then take the average and the quantiles of the 10,000 curves (that is, at each point t of the discretization grid, we take the average and the quantiles at 2.5% and 97.5% of the 10,000 occurrences) of $e^{-2\hat{\vartheta}_{2,n}(T_1-t)}\widehat{\sigma}^2_{n,t} + \widehat{\overline{\sigma}}^2_{n,t}$ and $e^{-2\bar{\vartheta}_{3,n}(T_1-t)}\widetilde{\sigma}^2_{3,n,t} + \widetilde{\overline{\sigma}}^2_{3,n,t}$. We should compare the plots to Figure 2.4 in the previous chapter, in which were the similar plots with no model errors.

In Figure 3.5, we see the plot of $e^{-2\hat{\vartheta}_{2,n}(T_1-t)}\widehat{\sigma}^2_{n,t} + \widehat{\overline{\sigma}}^2_{n,t}$ with $\beta = 0.8$. If we compare it with the plot of $e^{-2\bar{\vartheta}_{3,n}(T_1-t)}\widetilde{\sigma}^2_{3,n,t} + \widetilde{\overline{\sigma}}^2_{3,n,t}$ with the same value of β in Figure 3.6 and with the reference plot 2.4, we see that both give volatility estimates higher than the reference plot, and the upper quantiles are higher as well. The estimator with two processes is far less affected by errors than the one with three processes, as its upper quantile curve is lower than the one with three processes.
Now, if we take $\beta = 0.55$, both estimators overestimate volatility, but the estimator with 3 processes behaves better: it seems to be more robust to errors, performing in a worse manner than the other one when errors are low, but giving more accurate estimates when errors become higher. The plots are in Figures 3.7 and 3.8.
If we let n go to 1000 instead of 100, the quality of both estimators improves (compare with Figures 3.7 and 3.8). The plots are in Figures 3.9 and 3.10 in the appendices of the chapter. One can as well see in Figures 3.11 and 3.12, in the appendices, the plots with $\beta = 0.625$. This value is such that $\frac{\alpha}{2\alpha+1} = 2\beta - 1$, as we have $\alpha = 1/2$ with our specification of volatility; $\beta = 0.625$ is the lowest value such that the estimator with two processes reaches the optimal rate. Yet, the improvement that we get with three processes is noteworthy.

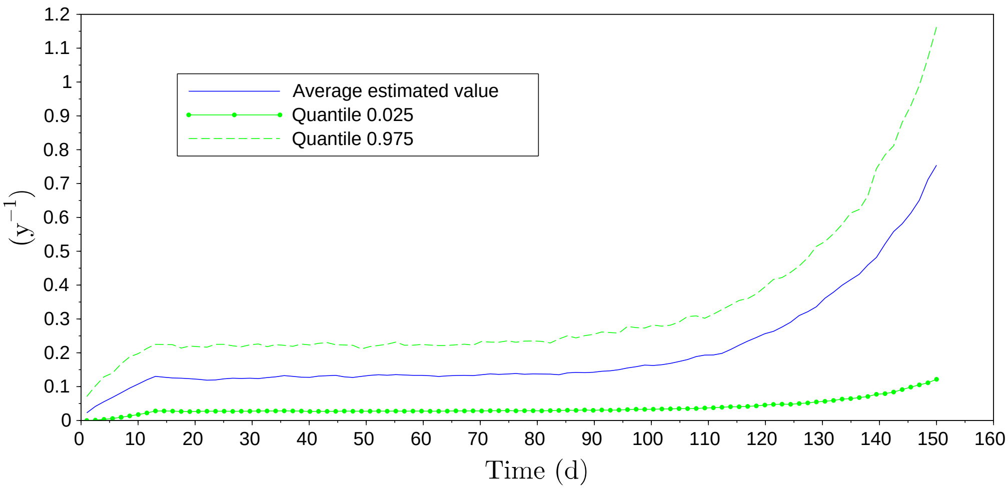

Figure 3.5 – Quantiles for the square of the equivalent volatility with 2 processes and $\beta = 0.8$

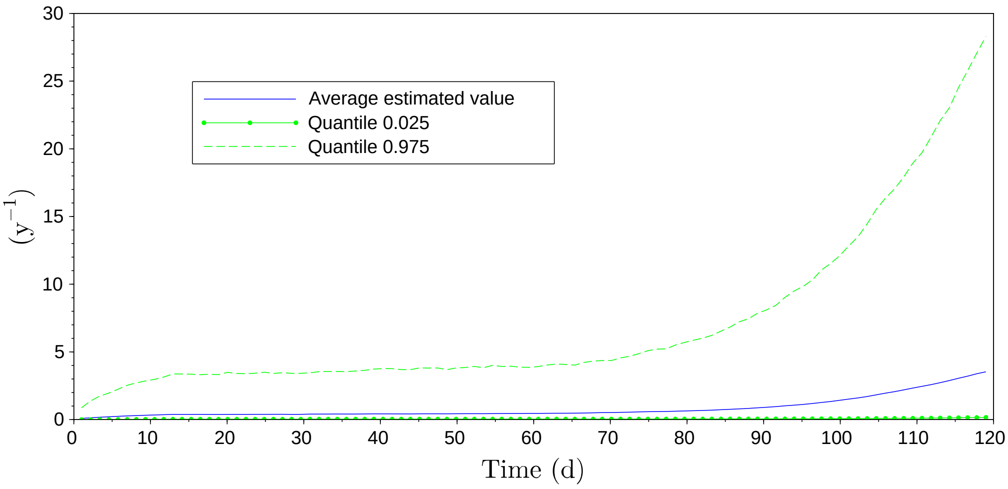

Figure 3.6 – Quantiles for the square of the equivalent volatility with 3 processes and $\beta = 0.8$

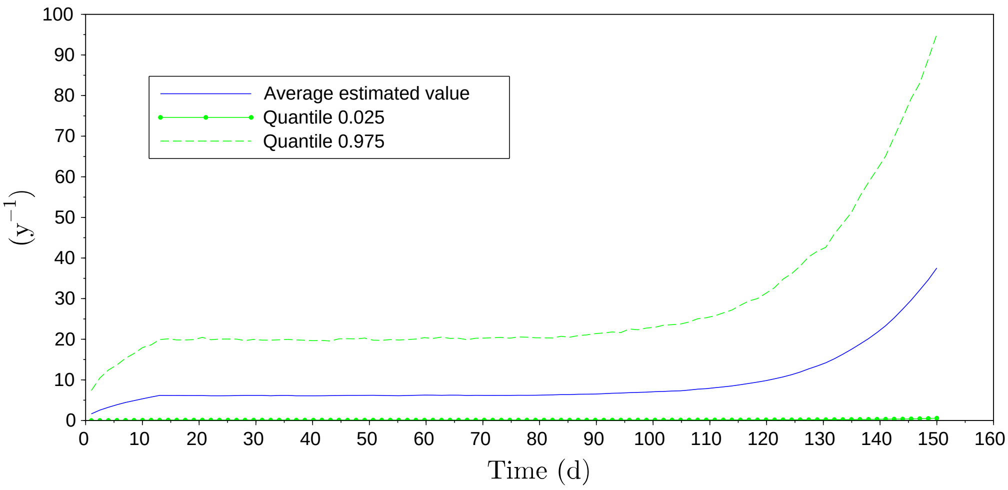

Figure 3.7 – Quantiles for the square of the equivalent volatility with 2 processes and $\beta = 0.55$

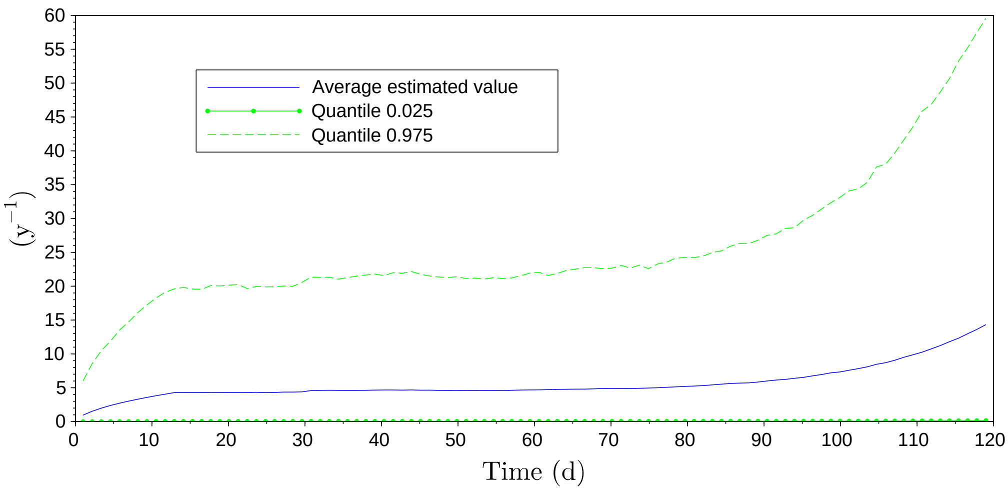

Figure 3.8 – Quantiles for the square of the equivalent volatility with 3 processes and $\beta = 0.55$

3.4 Proofs

3.4.1 Preliminaries: localization

Here we refer to the explanations in Section 2.4.1 ; while we have local boundedness only for the processes b, σ and $\overline{\sigma}$, we may use localization to lead the proofs under a boundedness assumption, which can be removed afterwards.
Thus, in all proofs we shall do as if b were bounded by some constant $M_b > 0$ and the two processes σ and $\overline{\sigma}$ were bounded by $M_\Sigma > 0$.

3.4.2 Proof of Theorem 3.1

Step 1 Let us prove the consistency of the estimator. We adopt the following notation, consistent with the one of the previous chapter: let $\mathfrak{e}_{\ell,k}(\vartheta)$ denote $e^{-\vartheta T_k} - e^{-\vartheta T_\ell}$, $\Delta_i^n \epsilon^j = \epsilon_i^j - \epsilon_{i-1}^j$,

$$\zeta_i^n = (\Delta_i^n X^2)^2 - (\Delta_i^n X^1)^2, \tilde{\zeta}_i^n = (\Delta_i^n Y^2)^2 - (\Delta_i^n Y^1)^2$$

and

$$\xi_i^n = (\Delta_i^n X^2 - \Delta_i^n X^1)^2, \tilde{\xi}_i^n = (\Delta_i^n Y^2 - \Delta_i^n Y^1)^2.$$

Recall that in Section 2.4.2, we proved that

$$\Psi_{T_1,T_2}^n = \frac{\sum_{i=1}^n \xi_i^n}{\sum_{i=1}^n \zeta_i^n} \to \psi_{T_1,T_2}(\vartheta)$$

in probability. Now,

$$\widetilde{\Psi}_{T_1,T_2}^n = \frac{\sum_{i=1}^n \tilde{\xi}_i^n}{\sum_{i=1}^n \tilde{\zeta}_i^n},$$

and because $\Delta_i^n Y^j = \Delta_i^n X^j + \kappa_j^n \Delta_i^n \epsilon^j$, $j = 1, 2$, we have

$$\tilde{\zeta}_i^n = \zeta_i^n + c_{1,i}^n + c_{2,i}^n \text{ and } \tilde{\xi}_i^n = \xi_i^n + c_{3,i}^n + c_{4,i}^n,$$

with

$$\begin{aligned} c_{1,i}^n &= 2\kappa_2^n \Delta_i^n \epsilon^2 \Delta_i^n X^2 - 2\kappa_1^n \Delta_i^n \epsilon^1 \Delta_i^n X^1, & c_{2,i}^n &= \left(\kappa_2^n \Delta_i^n \epsilon^2\right)^2 - \left(\kappa_1^n \Delta_i^n \epsilon^1\right)^2, \\ c_{3,i}^n &= 2\left(\kappa_2^n \Delta_i^n \epsilon^2 - \kappa_1^n \Delta_i^n \epsilon^1\right)\left(\Delta_i^n X^2 - \Delta_i^n X^1\right), & c_{4,i}^n &= \left(\kappa_2^n \Delta_i^n \epsilon^2 - \kappa_1^n \Delta_i^n \epsilon^1\right)^2. \end{aligned}$$

Now we prove that $\sum_{i=1}^n c_{k,i}^n \to 0$ in probability, $k = 1, \ldots, 4$.
Let $j, \ell = 1, 2$, and let $\gamma \geq 0$. By Tchebychev inequality, for all $\varepsilon > 0$,

$$\begin{aligned}
&\mathbb{P}\Big(\Big|n^\gamma \sum_{i=1}^n \kappa_j^n \Delta_i^n \epsilon^j \Delta_i^n X^\ell\Big| > \varepsilon\Big) \\
&\leq \frac{1}{\varepsilon^2}\mathbb{E}\Big(\Big(\sum_{i=1}^n n^\gamma \kappa_j^n \Delta_i^n \epsilon^j \Delta_i^n X^\ell\Big)^2\Big) \\
&= \frac{n^{2\gamma}(\kappa_j^n)^2}{\varepsilon^2}\Big(\sum_{i=1}^n \mathbb{E}\big(\big(\Delta_i^n \epsilon^j \Delta_i^n X^\ell\big)^2\big) + 2\sum_{1\leq i<k\leq n} \mathbb{E}\big(\Delta_i^n \epsilon^j \Delta_i^n X^\ell \Delta_k^n \epsilon^j \Delta_k^n X^\ell\big)\Big) \\
&= \frac{n^{2\gamma}(\kappa_j^n)^2}{\varepsilon^2}\Big(\sum_{i=1}^n \mathbb{E}\big(\big(\Delta_i^n \epsilon^j\big)^2\big)\mathbb{E}\big(\big(\Delta_i^n X^\ell\big)^2\big) + 2\sum_{i=1}^{n-1} \mathbb{E}\big(\Delta_i^n \epsilon^j \Delta_{i+1}^n \epsilon^j\big)\mathbb{E}\big(\Delta_i^n X^\ell \Delta_{i+1}^n X^\ell\big)\Big) \\
&= \frac{n^{2\gamma}(\kappa_j^n)^2}{\varepsilon^2}\Big(\sum_{i=1}^n 2\mathbb{E}\big(\big(\Delta_i^n X^\ell\big)^2\big) + 2\sum_{i=1}^{n-1} \mathbb{E}\big(\Delta_i^n X^\ell \Delta_{i+1}^n X^\ell\big)\Big),
\end{aligned}$$

where we have used independence of X and error terms. As, for $i = 1, \ldots, n$, $\mathbb{E}\big(\big(\Delta_i^n X^\ell\big)^2\big)$ and $\mathbb{E}\big(\big|\Delta_i^n X^\ell \Delta_{i+1}^n X^\ell\big|\big)$ are of order Δ_n, the whole term is $O(\Delta_n^{2(\beta-\gamma)})$ as $n \to \infty$. As this is true for all $\varepsilon > 0$ as well as for all j and ℓ, we have that as soon as $\gamma < \beta$, $n^\gamma \sum_{i=1}^n c_{1,i}^n$ and $n^\gamma \sum_{i=1}^n c_{3,i}^n$ converge to 0 in probability, as $n \to \infty$. Take $\gamma = 0$ to get the expected result.

At this point, we introduce some notation. For all t, let $\mathcal{G}_t$ be the smallest σ-algebra containing $\mathcal{F}_t$ and such that all random variables ϵ_i^j, for $t_i \leq t$ and $j = 1, \ldots, d$, are $\mathcal{G}_t$-measurable. Let $\mathcal{H}_t = \mathcal{G}_t \vee \mathcal{X}$, $\mathcal{X}$ being the σ-algebra generated by the processes X, σ, $\overline{\sigma}$ and b. To alleviate notation, we shall write $\mathcal{G}_i$ and $\mathcal{H}_i$, instead of, respectively, $\mathcal{G}_{t_i}$ and $\mathcal{H}_{t_i}$.

To care for $c_{2,i}^n$ and $c_{4,i}^n$, see that

$$\begin{aligned}
\mathbb{E}\big(|c_{2,i}^n|\big|\mathcal{G}_{i-1}\big) &\leq (\kappa_2^n)^2\mathbb{E}\big((\Delta_i^n \epsilon^2)^2\big|\mathcal{G}_{i-1}\big) + (\kappa_1^n)^2\mathbb{E}\big((\Delta_i^n \epsilon^1)^2\big|\mathcal{G}_{i-1}\big) \\
&= (\kappa_2^n)^2\big(1 + (\epsilon_{i-1}^2)^2\big) + (\kappa_1^n)^2\big(1 + (\epsilon_{i-1}^1)^2\big),
\end{aligned}$$

so that $\sum_{i=1}^n \mathbb{E}\big(|c_{2,i}^n|\big|\mathcal{F}_{i-1}\big) = O_{\mathbb{P}}(\Delta_n^{2\beta-1})$; the sum converges to 0 in probability, because we have $\beta > 1/2$. On the same way,

$$\mathbb{E}\big(|c_{4,i}^n|\big|\mathcal{G}_{i-1}\big) \leq 2\big((\kappa_2^n)^2\mathbb{E}\big((\Delta_i^n \epsilon^2)^2\big|\mathcal{G}_{i-1}\big) + (\kappa_1^n)^2\mathbb{E}\big((\Delta_i^n \epsilon^1)^2\big|\mathcal{G}_{i-1}\big)\big)$$

using the convexity inequality $(a + b)^2 \leq 2(a^2 + b^2)$, so that we may conclude that the sum $\sum_{i=1}^n \mathbb{E}\big(|c_{4,i}^n|\big|\mathcal{F}_{i-1}\big)$ converges to 0 in probability too. Those two sums converging to 0 correspond to a stronger version of condition (3.40) of Lemma 3.4 in [49], applied to $(c_{2,i}^n)_i$ and $(c_{4,i}^n)_i$, which proves that $\sum_{i=1}^n c_{2,i}^n$ and $\sum_{i=1}^n c_{4,i}^n$ converge to 0 in probability, as $n \to \infty$.

We have also proved that $\sum_{i=1}^n \zeta_i^n$ and $\sum_{i=1}^n \tilde{\zeta}_i^n$ have the same (almost-surely positive) limit in probability, and $\sum_{i=1}^n \xi_i^n$ and $\sum_{i=1}^n \tilde{\xi}_i^n$ have the same limit in probability too. Then

$$\widetilde{\Psi}_{T_1,T_2}^n = \frac{\sum_{i=1}^n \tilde{\xi}_i^n}{\sum_{i=1}^n \tilde{\zeta}_i^n} \to \psi_{T_1,T_2}(\vartheta)$$

in probability. Thus,

$$\psi_{T_1,T_2}\big(\hat{\vartheta}_{2,n}\big) \to \psi_{T_1,T_2}(\vartheta)$$

in probability on the event $\{\widetilde{\Psi}_{T_1,T_2} \in (-1,1)\}$, of which asymptotic probability is 1. As ψ_{T_1,T_2} is invertible, we get that $\hat{\vartheta}_{2,n}$ is a consistent estimator of ϑ.

Step 2 Now we establish the limit behaviour of the sequence of estimators $\hat{\vartheta}_{2,n}$. Let $\gamma > 0$. Write

$$\Delta_n^{-\gamma}\big(\widetilde{\Psi}_{T_1,T_2}^n - \psi_{T_1,T_2}(\vartheta)\big) = \Delta_n^{-\gamma}\frac{\sum_{i=1}^n \tilde{\xi}_i^n - \psi_{T_1,T_2}(\vartheta)\sum_{i=1}^n \tilde{\zeta}_i^n}{\sum_{i=1}^n \tilde{\zeta}_i^n},$$

and

$$\begin{aligned}\tilde{\xi}_i^n - \psi_{T_1,T_2}(\vartheta)\tilde{\zeta}_i^n &= \xi_i^n - \psi_{T_1,T_2}(\vartheta)\zeta_i^n \\ &\quad + c_{1,i}^n - \psi_{T_1,T_2}(\vartheta)c_{3,i}^n \\ &\quad + \big(1-\psi_{T_1,T_2}(\vartheta)\big)(\kappa_2^n)^2(\Delta_i^n\epsilon^2)^2 \\ &\quad - \big(1+\psi_{T_1,T_2}(\vartheta)\big)(\kappa_1^n)^2(\Delta_i^n\epsilon^1)^2 \\ &\quad + 2\psi_{T_1,T_2}(\vartheta)\kappa_1^n\kappa_2^n\Delta_i^n\epsilon^1\Delta_i^n\epsilon^2. \end{aligned} \tag{3.3}$$

In Step 2 of the proof in Section 2.4.2, we proved that

$$\Delta_n^{-\gamma}\sum_{i=1}^{\lfloor t\Delta_n^{-1}\rfloor}(\xi_i^n - \psi_{T_1,T_2}(\vartheta)\zeta_i^n) \to 2\psi_{T_1,T_2}(\vartheta)(e^{-\vartheta T_2} - e^{-\vartheta T_1})\chi(t)$$

stably in law, for $\gamma = 1/2$. Let us examine the other terms in the RHS of (3.3).
First, from Step 1 of the current proof, we have that $n^\gamma\sum_{i=1}^n c_{1,i}^n$ and $n^\gamma\sum_{i=1}^n c_{3,i}^n$ converge to 0 in probability, as $n\to\infty$, as soon as $\beta > \gamma$. So do $\Delta_n^{-\gamma}\sum_{i=1}^n c_{1,i}^n$ and $\Delta_n^{-\gamma}\psi_{T_1,T_2}(\vartheta)\sum_{i=1}^n c_{3,i}^n$. Then,

$$(\kappa_2^n)^2(\Delta_i^n\epsilon^2)^2 = (\kappa_2^n)^2((\epsilon_i^2)^2 + (\epsilon_{i-1}^2)^2 - 2\epsilon_i^2\epsilon_{i-1}^2),$$

and thus

$$\Delta_n^{1-2\beta}\sum_{i=1}^n(\kappa_2^n)^2(\Delta_i^n\epsilon^2)^2 \to 2T^{1-2\beta}\iota_2^2$$

in probability, using the classic law of large numbers for the sums of iid variables $\sum_{i=1}^n(\epsilon_i^2)^2$, $\sum_{i=1}^n(\epsilon_{i-1}^2)^2$, $\sum_{i=1,i\text{ odd}}^n\epsilon_i^2\epsilon_{i-1}^2$ and $\sum_{i=1,i\text{ even}}^n\epsilon_i^2\epsilon_{i-1}^2$ together with Assumption 3.1. Using the same method, we get that

$$\Delta_n^{1-2\beta}\sum_{i=1}^n\Big(\big(1-\psi_{T_1,T_2}(\vartheta)\big)(\kappa_2^n)^2(\Delta_i^n\epsilon^2)^2 - \big(1+\psi_{T_1,T_2}(\vartheta)\big)(\kappa_1^n)^2(\Delta_i^n\epsilon^1)^2\Big)$$

converges in probability to

$$2T^{1-2\beta}\Big(\big(1-\psi_{T_1,T_2}(\vartheta)\big)\iota_2^2 - \big(1+\psi_{T_1,T_2}(\vartheta)\big)\iota_1^2\Big).$$

Finally, with the splitting

$$\Delta_i^n \epsilon^1 \Delta_i^n \epsilon^2 = \epsilon_i^1 \epsilon_i^2 - \epsilon_{i-1}^1 \epsilon_i^2 - \epsilon_i^1 \epsilon_{i-1}^2 + \epsilon_{i-1}^1 \epsilon_{i-1}^2,$$

it becomes clear that the sum

$$\Delta_n^{-\gamma} \sum_{i=1}^n 2\psi_{T_1,T_2}(\vartheta) \kappa_1^n \kappa_2^n \Delta_i^n \epsilon^1 \Delta_i^n \epsilon^2$$

can be decomposed as the sum of sums of iid random variables, which are all $O_{\mathbb{P}}(n^{\gamma-2\beta+1/2})$ due to the classic CLT and Assumption 3.1.

In Step 1 of the current proof, we proved that $\sum_{i=1}^n \tilde{\zeta}_i^n \to \left(e^{-2\vartheta T_2} - e^{-2\vartheta T_1}\right) \int_0^T e^{2\vartheta t} \sigma_t^2 dt$ in probability, so that

— if $1/2 < \beta < 3/4$, we set $\gamma = 2\beta - 1$, and using the splitting (3.3) and the above results of convergence, we have that

$$\Delta_n^{1-2\beta}\left(\tilde{\Psi}_{T_1,T_2}^n - \psi_{T_1,T_2}(\vartheta)\right) \to m_{\vartheta,\beta} := \frac{2T^{1-2\beta}\left(\left(1-\psi_{T_1,T_2}(\vartheta)\right)\iota_2^2 - \left(1+\psi_{T_1,T_2}(\vartheta)\right)\iota_1^2\right)}{\left(e^{-2\vartheta T_2} - e^{-2\vartheta T_1}\right) \int_0^T e^{2\vartheta t} \sigma_t^2 dt}$$

in probability.

— if $3/4 < \beta$, we set $\gamma = 1/2$, and

$$\Delta_n^{-1/2}\left(\tilde{\Psi}_{T_1,T_2}^n - \psi_{T_1,T_2}(\vartheta)\right) \to \frac{2\psi_{T_1,T_2}(\vartheta)(e^{-\vartheta T_2} - e^{-\vartheta T_1})\chi(T)}{\left(e^{-2\vartheta T_2} - e^{-2\vartheta T_1}\right) \int_0^T e^{2\vartheta t} \sigma_t^2 dt}$$

in distribution; the limiting law is, conditionally to $\mathcal{F}$, centered Gaussian with conditional variance $v_\vartheta(\sigma, \overline{\sigma}) = 4\frac{(e^{-\vartheta T_2} - e^{-\vartheta T_1})^2 \int_0^T e^{2\vartheta t} \sigma_t^2 \overline{\sigma}_t^2 dB_t}{(e^{-\vartheta T_2} + e^{-\vartheta T_1})^4 \left(\int_0^T e^{2\vartheta t}\sigma_t^2 dt\right)^2}$.

— if $\beta = 3/4$, we set $\gamma = 1/2$, and

$$\Delta_n^{-1/2}\left(\tilde{\Psi}_{T_1,T_2}^n - \psi_{T_1,T_2}(\vartheta)\right) \to \frac{2\psi_{T_1,T_2}(\vartheta)(e^{-\vartheta T_2} - e^{-\vartheta T_1})\chi(T)}{\left(e^{-2\vartheta T_2} - e^{-2\vartheta T_1}\right) \int_0^T e^{2\vartheta t} \sigma_t^2 dt} + m_{\vartheta,\beta}$$

in distribution; the limiting law is, conditionally to $\mathcal{F}$, Gaussian with conditional mean $m_{\vartheta,\beta}$ and conditional variance $v_\vartheta(\sigma, \overline{\sigma})$.

Step 3 As we did in Step 3 of the proof in Section 2.4.2, we have, on the event $\{\widetilde{\Psi}_{T_1,T_2}^n \in (-1,0)\}$, which has asymptotically probability 1,

$$\Delta_n^{-\gamma}(\hat{\vartheta}_{2,n} - \vartheta) = \Delta_n^{-\gamma}(\widetilde{\Psi}_{T_1,T_2}^n - \psi_{T_1,T_2}(\vartheta))\partial_\vartheta \psi_{T,1,T_2}^{-1}(\widetilde{Z}_n),$$

for some $\widetilde{Z}_n$ which converges to $\psi_{T_1,T_2}(\vartheta)$ by Step 1. Then we conclude, as

$$\partial_\vartheta \psi_{T,1,T_2}^{-1}(\psi_{T_1,T_2}(\vartheta)) m_{\vartheta,\beta} = M_{\vartheta,\beta} \text{ and } \left(\partial_\vartheta \psi_{T,1,T_2}^{-1}(\psi_{T_1,T_2}(\vartheta))\right)^2 v_\vartheta(\sigma,\overline{\sigma}) = V_\vartheta(\sigma,\overline{\sigma}).$$

3.4.3 Proof of Theorem 3.2

Step 1 First we prove that the estimator is consistent. In Section 3.4.2, we proved that

$$\sum_{i=1}^{n}\big(\Delta_i^n(Y^2-Y^1)\big)^2=\sum_{i=1}^{n}\tilde{\xi}_i^n\to\mathfrak{e}_{1,2}(\vartheta)^2\int_0^T e^{2\vartheta t}\sigma_t^2dt$$

in probability, and the same arguments show that

$$\sum_{i=1}^{n}\big(\Delta_i^n(Y^j-Y^{j-1})\big)^2\to\mathfrak{e}_{j-1,j}(\vartheta)^2\int_0^T e^{2\vartheta t}\sigma_t^2dt$$

in probability too, for $j=2,\dots,d$. It is thus clear that

$$\widetilde{\Psi}_{T_{1..d}}^n\to\psi_{T_{1..d}}(\vartheta)$$

in probability, which means that $\psi_{T_{1..d}}(\hat{\vartheta}_{d,n})$ converges to $\psi_{T_{1..d}}(\vartheta)$ in probability, on the event $\big\{\widetilde{\Psi}_{T_{1..d}}^n\in\big(0,\sum_{j=3}^d\big(\frac{T_j-T_{j-1}}{T_2-T_1}\big)^2\big)\big\}$. Asymptotically, the probability of this event is 1, and $\psi_{T_{1..d}}$ is invertible because it is decreasing; this shows that $\hat{\vartheta}_{d,n}\to\vartheta$ in probability.

Step 2 Now we look at the asymptotic behaviour of the sequence of estimators $\hat{\vartheta}_{d,n}$. To do so, write

$$\widetilde{\Psi}_{T_{1..d}}^n-\psi_{T_{1..d}}(\vartheta)=\frac{\sum_{i=1}^n\tilde{\eta}_i^n}{\sum_{i=1}^n\big(\Delta_i^n(Y^2-Y^1)\big)^2},$$

with

$$\begin{aligned}\tilde{\eta}_i^n&=\sum_{j=3}^d\left(\big(\Delta_i^n(Y^j-Y^{j-1})\big)^2-\frac{\mathfrak{e}_{j-1,j}(\vartheta)^2}{\mathfrak{e}_{1,2}(\vartheta)^2}\big(\Delta_i^n(Y^2-Y^1)\big)^2\right)\\&=\sum_{j=3}^d(\eta_j)_i^n+(c_j)_{3,i}^n-\frac{\mathfrak{e}_{j-1,j}(\vartheta)^2}{\mathfrak{e}_{1,2}(\vartheta)^2}(c_2)_{3,i}^n+(c_j)_{4,i}^n-\frac{\mathfrak{e}_{j-1,j}(\vartheta)^2}{\mathfrak{e}_{1,2}(\vartheta)^2}(c_2)_{4,i}^n,\end{aligned}$$

the notation being inspired from the previous proof, but with one more index, j. We thus let

$$\begin{aligned}(\eta_j)_i^n=&\big(\Delta_i^n(X^j-X^{j-1})\big)^2-\frac{\mathfrak{e}_{j-1,j}(\vartheta)^2}{\mathfrak{e}_{1,2}(\vartheta)^2}\big(\Delta_i^n(X^2-X^1)\big)^2,\\(c_j)_{3,i}^n=&(\kappa_j^n\Delta_i^n\epsilon^j-\kappa_{j-1}^n\Delta_i^n\epsilon^{j-1})\big(\Delta_i^n(X^j-X^{j-1})\big),\\(c_j)_{4,i}^n=&(\kappa_j^n\Delta_i^n\epsilon^j-\kappa_{j-1}^n\Delta_i^n\epsilon^{j-1})^2.\end{aligned}$$

In Step 2 of the proof in Section 2.4.2, we proved that

$$\Delta_n^{-1}\sum_{i=1}^n(\eta_3)_i^n\to\int_0^T\mu_\vartheta^3(b_t)dt+2\int_0^T\lambda_\vartheta^3(b_t)e^{\vartheta t}\sigma_t dB_t,$$

in probability, where $\mu^j_\vartheta(b_t) = (b^j_t - b^{j-1}_t)^2 - \frac{\mathfrak{e}_{j-1,j}(\vartheta)^2}{\mathfrak{e}_{1,2}(\vartheta)^2}(b^2_t - b^1_t)^2$ and $\lambda^j_\vartheta(b_t) = \frac{\mathfrak{e}_{j-1,j}(\vartheta)}{\mathfrak{e}_{1,2}(\vartheta)}\big((b^j_t - b^{j-1}_t)\mathfrak{e}_{1,2}(\vartheta) - (b^2_t - b^1_t)\mathfrak{e}_{j-1,j}(\vartheta)\big)$. Referring to this proof, we see that the same method may be applied for $j > 3$, so that we get

$$\Delta_n^{-1} \sum_{i=1}^n \sum_{j=3}^d (\eta_j)^n_i \to \int_0^T \sum_{j=3}^d \mu^j_\vartheta(b_t)dt + 2\int_0^T \sum_{j=3}^d \lambda^j_\vartheta(b_t)e^{\vartheta t}\sigma_t dB_t \tag{3.4}$$

in probability.

In Step 2 of the proof in Section 3.4.2, we got

$$\Delta_n^{1-2\beta} \sum_{i=1}^n (c_2)^n_{4,i} \to 2T^{1-2\beta}(\iota_1^2 + \iota_2^2),$$

so that, using the same proof and summing over j then,

$$\Delta_n^{1-2\beta} \sum_{i=1}^n \sum_{j=3}^d \left((c_j)^n_{4,i} - \frac{\mathfrak{e}_{j-1,j}(\vartheta)^2}{\mathfrak{e}_{1,2}(\vartheta)^2}(c_2)^n_{4,i}\right) \to 2T^{1-2\beta}\sum_{j=3}^d \left((\iota_{j-1}^2 + \iota_j^2) - \frac{\mathfrak{e}_{j-1,j}(\vartheta)^2}{\mathfrak{e}_{1,2}(\vartheta)^2}(\iota_1^2 + \iota_2^2)\right) \tag{3.5}$$

in probability.

Finally, we saw in Section 3.4.2 that the sum $\Delta_n^{-\gamma}\sum_{i=1}^n \left(\sum_{j=3}^d (c_j)^n_{3,i} - \frac{\mathfrak{e}_{j-1,j}(\vartheta)^2}{\mathfrak{e}_{1,2}(\vartheta)^2}(c_2)^n_{3,i}\right)$ converges to 0 in probability if $\gamma < \beta$.

Step 3 Let us summarize using the convergences (3.4) and (3.5). If $1/2 < \beta < 1$, then

$$\Delta_n^{1-2\beta}\sum_{i=1}^n \tilde{\eta}^n_i \to 2T^{1-2\beta}\sum_{j=3}^d \left((\iota_{j-1}^2 + \iota_j^2) - \frac{\mathfrak{e}_{j-1,j}(\vartheta)^2}{\mathfrak{e}_{1,2}(\vartheta)^2}(\iota_1^2 + \iota_2^2)\right)$$

in probability. If $1 < \beta$, then

$$\Delta_n^{-1}\sum_{i=1}^n \tilde{\eta}^n_i \to \int_0^T \sum_{j=3}^d \mu^j_\vartheta(b_t)dt + 2\int_0^T \sum_{j=3}^d \lambda^j_\vartheta(b_t)e^{\vartheta t}\sigma_t dB_t$$

in probability.

Because

$$\widetilde{\Psi}^n_{T_{1..d}} - \psi_{T_{1..d}}(\vartheta) = \frac{\sum_{i=1}^n \tilde{\eta}^n_i}{\sum_{i=1}^n \left(\Delta^n_i(Y^2 - Y^1)\right)^2},$$

and

$$\sum_{i=1}^n \left(\Delta^n_i(Y^2 - Y^1)\right)^2 \to \mathfrak{e}_{1,2}(\vartheta)^2\int_0^T e^{2\vartheta t}\sigma_t^2 dt$$

in probability, that limit being defined on $(\Omega, \mathcal{F}, \mathbb{P})$, we derive the asymptotic behaviour of $\widetilde{\Psi}^n_{T_{1..d}} - \psi_{T_{1..d}}(\vartheta)$; if $1/2 < \beta < 1$, then

$$\Delta_n^{1-2\beta}\big(\widetilde{\Psi}^n_{T_{1..d}} - \psi_{T_{1..d}}(\vartheta)\big) \to \frac{2T^{1-2\beta}\sum_{j=3}^d \Big((\iota_{j-1}^2 + \iota_j^2) - \frac{\mathfrak{c}_{j-1,j}(\vartheta)^2}{\mathfrak{c}_{1,2}(\vartheta)^2}(\iota_1^2 + \iota_2^2)\Big)}{(e^{-\vartheta T_2} - e^{-\vartheta T_1})^2 \int_0^T e^{2\vartheta t}\sigma_t^2 dt}$$

in probability. If $1 < \beta$, then

$$\Delta_n^{-1}\big(\widetilde{\Psi}^n_{T_{1..d}} - \psi_{T_{1..d}}(\vartheta)\big) \to \frac{\int_0^T \sum_{j=3}^d \mu_\vartheta^j(b_t)dt + 2\int_0^T \sum_{j=3}^d \lambda_\vartheta^j(b_t)e^{\vartheta t}\sigma_t dB_t}{(e^{-\vartheta T_2} - e^{-\vartheta T_1})^2 \int_0^T e^{2\vartheta t}\sigma_t^2 dt}$$

in probability.

Step 4 On the event $\{\widetilde{\Psi}^n_{T_{1..d}} \in \big(0, \sum_{j=3}^d \big(\frac{T_j - T_{j-1}}{T_2 - T_1}\big)^2\big)\}$ (which has asymptotically probability 1),

$$\Delta_n^{-\gamma}(\hat{\vartheta}_{d,n} - \vartheta) = \Delta_n^{-\gamma}(\widetilde{\Psi}^n_{T_{1..d}} - \Psi_{T_{1..d}}(\vartheta))\partial_\vartheta \psi^{-1}_{T_{1..d}}(\widetilde{Z}_n),$$

for some $\widetilde{Z}_n$ which converges to $\psi_{T_{1..d}}(\vartheta)$ by Step 1. We can thus conclude the proof.

3.4.4 Proof of Theorem 3.3

Step 1 First we prove that the estimator $\bar{\vartheta}_{3,n}$ is consistent for ϑ. Let $1 \leq j < k \leq 3$, we have

$$\Delta_i^n Y^j \Delta_i^n Y^k = \Delta_i^n X^j \Delta_i^n X^k + a^j_{i,n}(k) + a^k_{i,n}(j) + b_{i,n}(j,k),$$

where

$$a^j_{i,n}(k) = \Delta_i^n X^j \kappa_k^n \Delta_i^n \epsilon^k \text{ and } b_{i,n}(j,k) = \kappa_j^n \Delta_i^n \epsilon^j \kappa_k^n \Delta_i^n \epsilon^k.$$

By usual convergence of quadratic variation,

$$\sum_{i=1}^n \Delta_i^n X^j \Delta_i^n X^k = \frac{1}{2}\Big(\sum_{i=1}^n (\Delta_i^n(X^j + X^k))^2 - \sum_{i=1}^n (\Delta_i^n X^j)^2 - \sum_{i=1}^n (\Delta_i^n X^k)^2\Big)$$

converges in probability to

$$\frac{1}{2}\Big(\int_0^T \big((e^{-\vartheta T_j} + e^{-\vartheta T_k})^2 e^{2\vartheta t}\sigma_t^2 + 4\overline{\sigma}_t^2\big)dt - \int_0^T (e^{-2\vartheta T_j}e^{2\vartheta t}\sigma_t^2 + \overline{\sigma}_t^2)dt - \int_0^T (e^{-2\vartheta T_k}e^{2\vartheta t}\sigma_t^2 + \overline{\sigma}_t^2)dt\Big),$$

that is, $\int_0^T (e^{-\vartheta(T_j + T_k)}e^{2\vartheta t}\sigma_t^2 + \overline{\sigma}_t^2)dt$. Then,

$$\begin{aligned}\mathbb{E}\Big(\Big(\sum_{i=1}^n a^j_{i,n}(k)\Big)^2\Big) &= \sum_{i=1}^n \mathbb{E}\big((a^j_{i,n}(k))^2\big) + 2\sum_{1\leq i<\ell\leq n} \mathbb{E}(a^j_{i,n}(k)a^j_{\ell,n}(k)) \\ &= \sum_{i=1}^n (\kappa_k^n)^2 \mathbb{E}((\Delta_i^n X^j)^2)\mathbb{E}((\Delta_i^n \epsilon^k)^2) - 2\sum_{i=1}^{n-1}(\kappa_k^n)^2 \mathbb{E}(\Delta_i^n X^j \Delta_{i+1}^n X^j)\end{aligned}$$

and those two sums converge to 0 as soon as $\beta > 0$, because of Assumption 3.1, and as $\mathbb{E}((\Delta_i^n X^j)^2)$ and $\mathbb{E}(\Delta_i^n X^j \Delta_{i+1}^n X^j)$ are of order Δ_n ; the sum $\sum_{i=1}^n a_{i,n}^j(k)$ converges to 0 in quadratic mean and thus in probability, for all indexes j and k. Finally,

$$
\begin{aligned}
\mathbb{E}\Big(\Big(\sum_{i=1}^n b_{i,n}(j,k)\Big)^2\Big) =& \sum_{i=1}^n \mathbb{E}\big((b_{i,n}(j,k))^2\big) + 2 \sum_{1\le i<\ell\le n} \mathbb{E}(b_{i,n}(j,k) b_{\ell,n}(j,k)) \\
=& (\kappa_j^n \kappa_k^n)^2 \Big(\sum_{i=1}^n \mathbb{E}\big(\Delta_i^n \epsilon^j \Delta_i^n \epsilon^k\big)^2 + 2\sum_{i=1}^{n-1} \mathbb{E}\big(\Delta_i^n \epsilon^j \Delta_{i+1}^n \epsilon^j\big) \mathbb{E}\big(\Delta_i^n \epsilon^k \Delta_{i+1}^n \epsilon^k\big)\Big)
\end{aligned}
$$

which is equivalent to $\frac{\iota_j^2 \iota_k^2}{n^{4\beta-1}}$ as $n \to \infty$, and therefore converges to 0 as soon as $\beta > 1/4$. Those results brought together show that

$$
\sum_{i=1}^n \Delta_i^n Y^j \Delta_i^n Y^k \to \int_0^T (e^{-\vartheta(T_j+T_k)} e^{2\vartheta t} \sigma_t^2 + \overline{\sigma}_t^2) dt
$$

in probability. It follows that

$$
\frac{\sum_{i=1}^n \Delta_i^n Y^1 (\Delta_i^n Y^2 - \Delta_i^n Y^3)}{\sum_{i=1}^n \Delta_i^n Y^2 (\Delta_i^n Y^1 - \Delta_i^n Y^3)} \to \frac{e^{-\vartheta T_1}(e^{-\vartheta T_2} - e^{-\vartheta T_3})}{e^{-\vartheta T_2}(e^{-\vartheta T_1} - e^{-\vartheta T_3})}
$$

because the limit in probability of $\sum_{i=1}^n \Delta_i^n Y^2 (\Delta_i^n Y^1 - \Delta_i^n Y^3)$ is non-zero almost-surely. We thus have

$$
\Phi_{T_1,T_2,T_3}^n \to \phi_{T_1,T_2,T_3}(\vartheta)
$$

in probability, hence $\bar{\vartheta}_{3,n} \to \vartheta$ in probability on the event $\big\{\Phi_{T_1,T_2,T_3}^n \in \big(\frac{T_3-T_2}{T_3-T_1}, 1\big)\big\}$. As this event has asymptotically probability 1, we get the convergence in probability of $\bar{\vartheta}_{3,n}$ towards ϑ.

Step 2 We now establish the limit law of $\Delta_n^{-1/2}\big(\Phi_{T_1,T_2,T_3}^n - \phi_{T_1,T_2,T_3}(\vartheta)\big)$. We have

$$
\Delta_n^{-1/2}\big(\Phi_{T_1,T_2,T_3}^n - \phi_{T_1,T_2,T_3}(\vartheta)\big) = \frac{\sum_{i=1}^n \chi_i^n}{e^{-\vartheta T_2}\big(e^{-\vartheta T_1} - e^{-\vartheta T_3}\big) \sum_{i=1}^n \Delta_i^n Y^2 (\Delta_i^n Y^1 - \Delta_i^n Y^3)},
$$

with

$$
\begin{aligned}
\chi_i^n =& \Delta_n^{-1/2}\big(e^{-\vartheta T_2} \mathfrak{e}_{3,1}(\vartheta) \Delta_i^n Y^1 (\Delta_i^n Y^2 - \Delta_i^n Y^3) - e^{-\vartheta T_1} \mathfrak{e}_{3,2}(\vartheta) \Delta_i^n Y^2 (\Delta_i^n Y^1 - \Delta_i^n Y^3)\big) \\
=& \Delta_n\big(e^{-\vartheta T_3} \mathfrak{e}_{2,1}(\vartheta) \Delta_i^n Y^1 \Delta_i^n Y^2 + e^{-\vartheta T_2} \mathfrak{e}_{1,3}(\vartheta) \Delta_i^n Y^1 \Delta_i^n Y^3 + e^{-\vartheta T_1} \mathfrak{e}_{3,2}(\vartheta) \Delta_i^n Y^2 \Delta_i^n Y^3\big).
\end{aligned}
$$

We shall work with Lemma 3.7 in [49], applied to variables χ_i^n. We therefore have to check conditions (3.43)–(3.46) of the lemma. To do so, we first observe that $\mathbb{E}(\chi_i^n | \mathcal{G}_{i-1})$ is equal to the expectation of

$$\begin{aligned}
&\Delta_n^{-1/2} e^{-\vartheta T_3} \mathfrak{e}_{2,1}(\vartheta)\Big(\int_{t_{i-1}}^{t_i} b_t^1 dt \int_{t_{i-1}}^{t_i} b_t^2 dt + \kappa_2^n \epsilon_{i-1}^2 \int_{t_{i-1}}^{t_i} b_t^1 dt + \kappa_1^n \epsilon_{i-1}^1 \int_{t_{i-1}}^{t_i} b_t^2 dt + \kappa_1^n \kappa_2^n \epsilon_{i-1}^1 \epsilon_{i-1}^2 \Big) \\
&+\Delta_n^{-1/2} e^{-\vartheta T_2} \mathfrak{e}_{1,3}(\vartheta)\Big(\int_{t_{i-1}}^{t_i} b_t^1 dt \int_{t_{i-1}}^{t_i} b_t^3 dt + \kappa_3^n \epsilon_{i-1}^3 \int_{t_{i-1}}^{t_i} b_t^1 dt + \kappa_1^n \epsilon_{i-1}^1 \int_{t_{i-1}}^{t_i} b_t^3 dt + \kappa_1^n \kappa_3^n \epsilon_{i-1}^1 \epsilon_{i-1}^3 \Big) \\
&+\Delta_n^{-1/2} e^{-\vartheta T_1} \mathfrak{e}_{3,2}(\vartheta)\Big(\int_{t_{i-1}}^{t_i} b_t^2 dt \int_{t_{i-1}}^{t_i} b_t^3 dt + \kappa_3^n \epsilon_{i-1}^3 \int_{t_{i-1}}^{t_i} b_t^2 dt + \kappa_2^n \epsilon_{i-1}^2 \int_{t_{i-1}}^{t_i} b_t^3 dt + \kappa_2^n \kappa_3^n \epsilon_{i-1}^2 \epsilon_{i-1}^3 \Big)
\end{aligned}$$

conditionally to $\mathcal{G}_{i-1}$.
Let $t \in [0,T]$ and $1 \leq j,k \leq 3, j \neq k$; we have $\mathbb{E}\Big(\Big| \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \Delta_n^{-1/2} \int_{t_{i-1}}^{t_i} b_t^j dt \int_{t_{i-1}}^{t_i} b_t^k dt \Big| \Big) \lesssim \Delta_n^{1/2}$, which converges to 0 in probability.
In the same way, $\mathbb{E}\Big(\Big| \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \Delta_n^{-1/2} \kappa_j^n \epsilon_{i-1}^j \int_{t_{i-1}}^{t_i} b_t^k dt \Big| \Big) \lesssim \Delta_n^{\beta-1/2}$, which goes to 0 too, as we assumed that $\beta > 1/2$.
Moreover, by the usual CLT, $\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \Delta_n^{1/2} \epsilon_{i-1}^j \epsilon_{i-1}^k$ converges in law to some distribution; then,

$$\mathbb{E}\Big(\Big| \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \Delta_n^{-1/2} \kappa_j^n \kappa_k^n \epsilon_{i-1}^j \epsilon_{i-1}^k \Big| \Big) = \Delta_n^{-1} \kappa_j^n \kappa_k^n \mathbb{E}\Big(\Big| \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \Delta_n^{1/2} \epsilon_{i-1}^j \epsilon_{i-1}^k \Big| \Big)$$

converges to 0 in probability, because $\Delta_n^{-1} \kappa_j^n \kappa_k^n \simeq \Delta_n^{2\beta-1}$, which goes to 0 as $n \to \infty$. Summing over j and k, it comes that

$$\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}\big(\big| \mathbb{E}(\chi_i^n | \mathcal{G}_{i-1}) \big| | \mathcal{G}_{i-1} \big) \to 0$$

in probability, which, by Lemma 3.4 in [49], ensures condition (3.43) with $A_t = 0$.

To care for condition (3.44), fix $t \in [0,T]$; as $\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \big(\mathbb{E}(\chi_i^n | \mathcal{G}_{i-1}) \big)^2 \to 0$ in probability, we are left with computing the limit in probability of $\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}\big((\chi_i^n)^2 \big| \mathcal{G}_{i-1} \big)$. To do so, in view of the definition of χ_i^n, we have to compute, for indexes $1 \leq j,k,m \leq 3, j \neq k, j \neq m$, the conditional expectation $\mathbb{E}\big(\Delta_n^{-1} (\Delta_i^n Y^j)^2 \Delta_i^n Y^k \Delta_i^n Y^m \big| \mathcal{G}_{i-1} \big)$.
Because $\beta > 1/2$, the terms including model errors will be negligible, and it is enough to compute $\mathbb{E}\big(\Delta_n^{-1} (\Delta_i^n X^j)^2 \Delta_i^n X^k \Delta_i^n X^m \big| \mathcal{G}_{i-1} \big)$, which is done using Itô formula.

We find that $\mathbb{E}\big((\Delta_i^n X^j)^2 \Delta_i^n X^k \Delta_i^n X^m \big| \mathcal{G}_{i-1}\big)$ is equal to

$$
\begin{aligned}
&\mathbb{E}\Big[\int_{t_{i-1}}^{t_i} \Big(\int_{t_{i-1}}^{t} e^{-\vartheta(T_k-s)}\sigma_s dB_s + \int_{t_{i-1}}^{t} \overline{\sigma}_s d\overline{B}_s\Big) \\
&\qquad \times \Big(\int_{t_{i-1}}^{t} e^{-\vartheta(T_m-s)}\sigma_s dB_s + \int_{t_{i-1}}^{t} \overline{\sigma}_s d\overline{B}_s\Big)\big(e^{-2\vartheta(T_j-t)}\sigma_t^2 + \overline{\sigma}_t^2\big)\Big|\mathcal{G}_{i-1}\Big] \\
+2&\mathbb{E}\Big[\int_{t_{i-1}}^{t_i} \Big(\int_{t_{i-1}}^{t} e^{-\vartheta(T_j-s)}\sigma_s dB_s + \int_{t_{i-1}}^{t} \overline{\sigma}_s d\overline{B}_s\Big) \\
&\qquad \times \Big(\int_{t_{i-1}}^{t} e^{-\vartheta(T_m-s)}\sigma_s dB_s + \int_{t_{i-1}}^{t} \overline{\sigma}_s d\overline{B}_s\Big)\big(e^{-\vartheta(T_j+T_k-2t)}\sigma_t^2 + \overline{\sigma}_t^2\big)\Big|\mathcal{G}_{i-1}\Big] \\
+2&\mathbb{E}\Big[\int_{t_{i-1}}^{t_i} \Big(\int_{t_{i-1}}^{t} e^{-\vartheta(T_j-s)}\sigma_s dB_s + \int_{t_{i-1}}^{t} \overline{\sigma}_s d\overline{B}_s\Big) \\
&\qquad \times \Big(\int_{t_{i-1}}^{t} e^{-\vartheta(T_k-s)}\sigma_s dB_s + \int_{t_{i-1}}^{t} \overline{\sigma}_s d\overline{B}_s\Big)\big(e^{-\vartheta(T_j+T_m-2t)}\sigma_t^2 + \overline{\sigma}_t^2\big)\Big|\mathcal{G}_{i-1}\Big] \\
+&\mathbb{E}\Big[\int_{t_{i-1}}^{t_i} \Big(\int_{t_{i-1}}^{t} e^{-\vartheta(T_j-s)}\sigma_s dB_s + \int_{t_{i-1}}^{t} \overline{\sigma}_s d\overline{B}_s\Big)^2\big(e^{-\vartheta(T_k+T_m-2t)}\sigma_t^2 + \overline{\sigma}_t^2\big)\Big|\mathcal{G}_{i-1}\Big]
\end{aligned}
$$

plus some terms including the drift process, which are of lower order. Then we use the technical lemma 3.6.2, which is stated and proved in Section 3.5.1, to get that for all $t \in [0,T]$, the sum $\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}\big(\Delta_n^{-1}(\Delta_i^n X^j)^2 \Delta_i^n X^k \Delta_i^n X^m \big| \mathcal{G}_{i-1}\big)$ converges in probability to

$$
\begin{aligned}
&3e^{-2\vartheta(T_k+T_m+2T_j)}\int_0^t e^{4\vartheta s}\sigma_s^4 ds + 3\int_0^t \overline{\sigma}_s^4 ds \\
&+(e^{-\vartheta(T_k+T_m)} + 2e^{-\vartheta(T_j+T_k)} + 2e^{-\vartheta(T_j+T_m)} + e^{-2\vartheta T_j})\int_0^t e^{2\vartheta s}\sigma_s^2\overline{\sigma}_s^2 ds
\end{aligned}
$$

as $n \to \infty$. As we have just said, $\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}\big(\Delta_n^{-1}(\Delta_i^n Y^j)^2 \Delta_i^n Y^k \Delta_i^n Y^m \big| \mathcal{G}_{i-1}\big)$ has the same limit. There remains to sum over j, k and m according to the definition of χ_i^n, and after some computations, we get that

$$
\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}\big((\chi_i^n)^2 \big| \mathcal{G}_{i-1}\big) \to C_t = (e^{-\vartheta T_1} - e^{-\vartheta T_2})^2(e^{-\vartheta T_2} - e^{-\vartheta T_3})^2(e^{-\vartheta T_3} - e^{-\vartheta T_1})^2\int_0^t e^{2\vartheta s}\sigma_s^2\overline{\sigma}_s^2 ds
$$

in probability. This is condition (3.44).

Now, using the convexity inequality $(a+b+c)^4 \leq 27(a^4+b^4+c^4)$ and the definition of χ_i^n, there is some constant K such that

$$
\begin{aligned}
&\mathbb{E}\big(|\chi_i^n|^4 \big| \mathcal{G}_{i-1}\big) \\
\leq &K\Delta_n^{-2}\big(\mathbb{E}\big(|\Delta_i^n Y^1 \Delta_i^n Y^2|^4 \big| \mathcal{G}_{i-1}\big) + \mathbb{E}\big(|\Delta_i^n Y^1 \Delta_i^n Y^3|^4 \big| \mathcal{G}_{i-1}\big) + \mathbb{E}\big(|\Delta_i^n Y^2 \Delta_i^n Y^3|^4 \big| \mathcal{G}_{i-1}\big)\big).
\end{aligned}
$$

In each expectation $\mathbb{E}\big(|\Delta_i^n Y^j \Delta_i^n Y^k|^4 \big| \mathcal{G}_{i-1}\big)$, the predominant term is $\mathbb{E}\big(|\Delta_i^n X^j \Delta_i^n X^k|^4 \big| \mathcal{G}_{i-1}\big)$, of which order is Δ_n^4. Therefore,

$$\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}\big(|\chi_i^n|^4 \big| \mathcal{G}_{i-1}\big) \lesssim \Delta_n,$$

which converges to 0 in probability as $n \to \infty$. This ensures condition (3.45).

To check condition (3.46), we have to show that

$$\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}\big(\chi_i^n \Delta_i^n N \big| \mathcal{G}_{i-1}\big) \to 0$$

in probability, whenever N is B, $\overline{B}$, or any bounded martingale orthogonal to $(B, \overline{B})$. For simplicity of presentation, we omit the drift and model error terms as they are lower-order ones and there is no difficulty related to them; we just need to use the fact that the drift process is continuous in probability, which is ensured by Assumption 2.1. Let $1 \le j, k \le 3, j \neq k$:

$$\mathbb{E}\big(\chi_i^n \Delta_i^n N \big| \mathcal{G}_{i-1}\big) = A_i^n + B_i^n + D_i^n,$$

where

$$\begin{aligned}
A_i^n =& \mathbb{E}\Big[\Delta_n^{-1/2}\Big(\sigma_{t_{i-1}} \int_{t_{i-1}}^{t_i} e^{-\vartheta(T_j - t)} dB_t + \overline{\sigma}_{t_{i-1}} \Delta_i^n \overline{B}\Big) \\
&\times \Big(\sigma_{t_{i-1}} \int_{t_{i-1}}^{t_i} e^{-\vartheta(T_k - t)} dB_t + \overline{\sigma}_{t_{i-1}} \Delta_i^n \overline{B}\Big) \Delta_i^n N \Big| \mathcal{G}_{i-1}\Big], \\
B_i^n =& \mathbb{E}\Big[\Delta_n^{-1/2}\Big(\int_{t_{i-1}}^{t_i} e^{-\vartheta(T_j - t)} (\sigma_t - \sigma_{t_{i-1}}) dB_t + \int_{t_{i-1}}^{t_i} (\overline{\sigma}_t - \overline{\sigma}_{t_{i-1}}) d\overline{B}_t\Big) \\
&\times \Big(\sigma_{t_{i-1}} \int_{t_{i-1}}^{t_i} e^{-\vartheta(T_k - t)} dB_t + \overline{\sigma}_{t_{i-1}} \Delta_i^n \overline{B}\Big) \Delta_i^n N \Big| \mathcal{G}_{i-1}\Big], \\
D_i^n =& \mathbb{E}\Big[\Delta_n^{-1/2}\Big(\int_{t_{i-1}}^{t_i} e^{-\vartheta(T_j - t)} \sigma_t dB_t + \int_{t_{i-1}}^{t_i} \overline{\sigma}_t d\overline{B}_t\Big) \\
&\times \Big(\int_{t_{i-1}}^{t_i} e^{-\vartheta(T_k - t)} (\sigma_t - \sigma_{t_{i-1}}) dB_t + \int_{t_{i-1}}^{t_i} (\overline{\sigma}_t - \overline{\sigma}_{t_{i-1}}) d\overline{B}_t\Big) \Delta_i^n N \Big| \mathcal{G}_{i-1}\Big].
\end{aligned}$$

Whenever N is B, $\overline{B}$ or a bounded martingale independent of $(B, \overline{B})$, the expectation in A_i^n is zero, as all integrals are Wiener integrals. Now, by Cauchy-Schwarz inequality,

$$\begin{aligned}
&\mathbb{E}\Big(\Big|\mathbb{E}\Big(\Delta_n^{-1/2} \int_{t_{i-1}}^{t_i} e^{-\vartheta(T_j - t)} (\sigma_t - \sigma_{t_{i-1}}) dB_t \overline{\sigma}_{t_{i-1}} \Delta_i^n \overline{B} \Delta_i^n N \Big| \mathcal{G}_{i-1}\Big)\Big|\Big) \\
\le& \Delta_n^{-1/2} \sqrt{\mathbb{E}\Big(\Big(\int_{t_{i-1}}^{t_i} e^{-\vartheta(T_j - t)} (\sigma_t - \sigma_{t_{i-1}}) dB_t\Big)^2\Big) \mathbb{E}\Big(\Big(\overline{\sigma}_{t_{i-1}} \Delta_i^n \overline{B} \Delta_i^n N\Big)^2\Big)}.
\end{aligned}$$

We have

$$\mathbb{E}\Big(\Big(\int_{t_{i-1}}^{t_i} e^{-\vartheta(T_j-t)}(\sigma_t-\sigma_{t_{i-1}})dB_t\Big)^2\Big) = \int_{t_{i-1}}^{t_i} e^{-2\vartheta(T_j-t)}\mathbb{E}\big((\sigma_t-\sigma_{t_{i-1}})^2\big)dt$$
$$\leq \int_{t_{i-1}}^{t_i} e^{-2\vartheta(T_j-t)}\mathbb{E}\Big(\frac{(\sigma_t^2-\sigma_{t_{i-1}}^2)^2}{(\sigma_t+\sigma_{t_{i-1}})^2}\Big)dt$$
$$\leq \frac{c\Delta_n^{1+2\alpha}}{4\tilde{c}^2},$$

and $\mathbb{E}\Big(\Big(\overline{\sigma}_{t_{i-1}}\Delta_i^n\overline{B}\Delta_i^n N\Big)^2\Big)$ is of order Δ_n^2 if $N=B$ or $N=\overline{B}$. If N is a bounded martingale orthogonal to $(B,\overline{B})$, then $\mathbb{E}\Big(\Big(\overline{\sigma}_{t_{i-1}}\Delta_i^n\overline{B}\Delta_i^n N\Big)^2\Big) \leq M_\Sigma^2\mathbb{E}\Big(\Big(\Delta_i^n\overline{B}\Big)^2\Big)\mathbb{E}\Big(\Big(\Delta_i^n N\Big)^2\Big) = M_\Sigma^2\Delta_n K$, for some constant K, because N is bounded. It comes that

$$\mathbb{E}\Big(\Big|\mathbb{E}\Big(\Delta_n^{-1/2}\int_{t_{i-1}}^{t_i} e^{-\vartheta(T_j-t)}(\sigma_t-\sigma_{t_{i-1}})dB_t\overline{\sigma}_{t_{i-1}}\Delta_i^n\overline{B}\Delta_i^n N\Big|\mathcal{G}_{i-1}\Big)\Big|\Big)$$

is at least of order $\Delta_n^{1/2+\alpha}$. We get the same bound for the other terms appearing in B_i^n and D_i^n, so that

$$\sum_{i=1}^{\lfloor t/\Delta_n\rfloor}(B_i^n+D_i^n)\to 0$$

in probability, as $\alpha>1/2$. By summation on j and k, we get condition (3.46). We may thus apply Lemma 3.7 in [49], and get, for $t=T$, that

$$\Delta_n^{-1/2}\sum_{i=1}^{n}\chi_i^n \to \mathcal{N}(0,C_T),$$

stably in law, where $\mathcal{N}(0,C_T)$ is a random variable defined on an extension of $(\Omega,\mathcal{F},\mathcal{P})$, which conditionally to $\mathcal{F}$ is Gaussian, centered with (conditional) variance C_T.

We showed in Step 1 that $\sum_{i=1}^n \Delta_i^n Y^2(\Delta_i^n Y^1-\Delta_i^n Y^3) \to e^{-\vartheta T_2}(e^{-\vartheta T_1}-e^{-\vartheta T_3})\int_0^T e^{2\vartheta t}\sigma_t^2 dt$ in probability, this limit being non-zero. Because the above convergence stands stably in law, we have

$$\Delta_n^{-1/2}\big(\Phi_{T_1,T_2,T_3}^n-\phi_{T_1,T_2,T_3}(\vartheta)\big)\to\mathcal{N}(0,v_3(T)),$$

stably in law, where $\mathcal{N}(0,v_3(T))$ is a random variable defined on an extension of $(\Omega,\mathcal{F},\mathcal{P})$, which conditionally to $\mathcal{F}$ is Gaussian, centered with (conditional) variance

$$v_3(T)=\frac{C_T}{e^{-4\vartheta T_2}(e^{-\vartheta T_1}-e^{-\vartheta T_3})^4\Big(\int_0^T e^{2\vartheta t}\sigma_t^2 dt\Big)^2}.$$

Step 3 There remains to establish the limit law of $\Delta_n^{-1/2}(\bar{\vartheta}_{3,n}-\vartheta)$. To do so, observe that

$$\Delta_n^{-1/2}(\bar{\vartheta}_{3,n}-\vartheta)=\Delta_n^{-1/2}\big(\Phi^n_{T_1,T_2,T_3}-\phi_{T_1,T_2,T_3}(\vartheta)\big)\partial_\vartheta\phi^{-1}_{T_1,T_2,T_3}(Z_n),$$

for some Z_n which, by Step 1, converges in probability to $\phi_{T_1,T_2,T_3}(\vartheta)$. We therefore conclude by noting that

$$\partial_\vartheta\phi_{T_1,T_2,T_3}(\vartheta)=\frac{e^{-\vartheta(T_1+T_2+T_3)}\big(T_1(e^{-\vartheta T_2}-e^{-\vartheta T_3})+T_2(e^{-\vartheta T_3}-e^{-\vartheta T_1})+T_3(e^{-\vartheta T_1}-e^{-\vartheta T_2})\big)}{e^{-2\vartheta T_2}\big(e^{-\vartheta T_1}-e^{-\vartheta T_3}\big)^2}.$$

3.4.5 Proof of Theorem 3.4

Let us look at the properties of the estimator $\widehat{\sigma}^2_{n,t}$. For ease of notation, we write $\hat{\vartheta}_{2,n}$ for $\max\{\hat{\vartheta}_{2,n},\varpi_n\}$ and set $t_i=i\Delta_n$ for $i=1,\dots,n$. We also define $K(t)=\mathbf{1}_{(0,1]}(t)$ and $K_h(t)=h^{-1}K(th^{-1})$ for $h>0$. We have

$$\widehat{\sigma}^2_{n,t}-\sigma^2_t=\frac{\sum_{i=1}^n K_{h_n}\big(t-t_{i-1}\big)\big((\Delta_i^nY^1)^2-(\Delta_i^nY^2)^2\big)}{e^{-2\hat{\vartheta}_{2,n}(T_1-t)}-e^{-2\hat{\vartheta}_{2,n}(T_2-t)}}-\sigma^2_t=I+II,$$

with

$$I=\Big(\frac{1}{e^{-2\hat{\vartheta}_{2,n}(T_1-t)}-e^{-2\hat{\vartheta}_{2,n}(T_2-t)}}-\frac{1}{e^{-2\vartheta(T_1-t)}-e^{-2\vartheta(T_2-t)}}\Big)\times\sum_{i=1}^n K_{h_n}\big(t-t_{i-1}\big)\big((\Delta_i^nY^1)^2-(\Delta_i^nY^2)^2\big)$$

and

$$II=\frac{\sum_{i=1}^n K_{h_n}\big(t-t_{i-1}\big)\big((\Delta_i^nY^1)^2-(\Delta_i^nY^2)^2\big)}{e^{-2\vartheta(T_1-t)}-e^{-2\vartheta(T_2-t)}}-\sigma^2_t.$$

The term I Since $(\Delta_i^nY^j)^2=(\Delta_i^nX^j)^2+2\kappa_1^n\Delta_i^n\epsilon^1\Delta_i^nX^j+(\kappa_1^n\Delta_i^n\epsilon^1)^2$, we have

$$(\Delta_i^nY^1)^2-(\Delta_i^nY^2)^2=(\Delta_i^nX^1)^2-(\Delta_i^nX^2)^2-c^n_{1,i}-c^n_{2,i},$$

where $c^n_{1,i}$ and $c^n_{2,i}$ are defined in Section 3.4.2. As $E(|c^n_{1,i}|)$ and $E(|c^n_{2,i}|)$ are respectively of order $\Delta_n^{\beta+1/2}$ and $\Delta_n^{2\beta}$, and $\mathbb{E}\big(\big|(\Delta_i^nX^1)^2-(\Delta_i^nX^2)^2\big|\big)$ is of order Δ_n,

$$\mathbb{E}\Big[\Big|\sum_{i=1}^n K_{h_n}(t-t_{i-1})\big((\Delta_i^nY^1)^2-(\Delta_i^nY^2)^2\big)\Big|\Big]\lesssim\sum_{i=1}^n K_{h_n}(t-t_{i-1})\Delta_n\lesssim 1$$

since $K_{h_n}(t-t_{i-1})$ is of order h_n^{-1} for a number of terms that are at most of order $\Delta_n^{-1}h_n$. Therefore $\sum_{i=1}^n K_{h_n}\big(t-t_{i-1}\big)\big((\Delta_i^nY^1)^2-(\Delta_i^nY^2)^2\big)$ is bounded in expectation hence tight. The order of I is thus dictated by

$$\frac{1}{e^{-2\hat{\vartheta}_{2,n}(T_1-t)}-e^{-2\hat{\vartheta}_{2,n}(T_2-t)}}-\frac{1}{e^{-2\vartheta(T_1-t)}-e^{-2\vartheta(T_2-t)}}$$

which, by Theorem 3.1, is of order $\Delta_n^{1/2\wedge(2\beta-1)}$.

The term II The term II further splits into $II = (e^{-2\vartheta(T_1-t)} - e^{-2\vartheta(T_2-t)})^{-1}\big(B_n(t) + V_n(t)\big)$, having

$$V_n(t) = \sum_{i=1}^{n} K_{h_n}(t - t_{i-1})\big((\Delta_i^n Y^1)^2 - (\Delta_i^n Y^2)^2 - \mathbb{E}\big[(\Delta_i^n Y^1)^2 - (\Delta_i^n Y^2)^2 \big| \mathcal{G}_{i-1}\big]\big)$$

and

$$B_n(t) = \sum_{i=1}^{n} \mathbb{E}\big[K_{h_n}(t - t_{i-1})\big((\Delta_i^n Y^1)^2 - (\Delta_i^n Y^2)^2\big)\big|\mathcal{G}_{i-1}\big] - \big(e^{-2\vartheta(T_1-t)} - e^{-2\vartheta(T_2-t)}\big)\sigma_t^2.$$

Bounding the variance term

We first prove an upper bound for $\mathbb{E}(V_n(t)^2)$ uniformly in $t \in [h_n, T]$. In Section 2.4.3, we proved that

$$\sup_{t\in[h_n,T]} \mathbb{E}\Big[\Big(\sum_{i=1}^{n} K_{h_n}(t - t_{i-1})\big((\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2 - \mathbb{E}\big[(\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2 \big| \mathcal{F}_{i-1}\big]\big)\Big)^2\Big]$$
$$\lesssim \Delta_n h_n^{-1},$$

and while examining the term I, we wrote the decomposition

$$(\Delta_i^n Y^1)^2 - (\Delta_i^n Y^2)^2 = (\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2 - c_{1,i}^n - c_{2,i}^n.$$

As $\mathbb{E}\big(|c_{1,i}^n + c_{2,i}^n|^2\big)$ is a term of lower order than $\mathbb{E}\big(\big|(\Delta_i^n X^1)^2 - (\Delta_i^n X^2)^2\big|^2\big)$, we obtain $\sup_{t\in[h_n,T]} \mathbb{E}\big[(V_n(t))^2\big] \lesssim \Delta_n h_n^{-1}$.

Bounding the bias term

In order to get an upper bound for the bias term $\sup_{t\in[h_n,T]} \mathbb{E}\big((B_n(t))^2\big)$, we use the decomposition

$$B_n(t) = \big(e^{-2\vartheta(T_1-t)} - e^{-2\vartheta(T_2-t)}\big)(B_I(t) + B_{II}(t)),$$

where

$$B_I(t) = \int_0^T h_n^{-1} K\Big(\frac{t-u}{h_n}\Big) e^{-2\vartheta(t-u)} \sigma_u^2 du - \sigma_t^2$$

and

$$B_{II}(t) = \frac{\sum_{i=1}^{n} \mathbb{E}\Big(h_n^{-1} K\Big(\frac{t-t_{i-1}}{h_n}\Big)\big((\Delta_i^n Y^1)^2 - (\Delta_i^n Y^2)^2\big)\Big|\mathcal{G}_{i-1}\Big)}{e^{-2\vartheta(T_1-t)} - e^{-2\vartheta(T_2-t)}}$$
$$- \int_0^T h_n^{-1} K\Big(\frac{t-u}{h_n}\Big) e^{-2\vartheta(t-u)} \sigma_u^2 du.$$

The term B_I has been treated in Section 2.4.3 : we got

$$\sup_{t\in[h_n,T]} \mathbb{E}\big((B_I(t))^2\big) \lesssim h_n^{2\alpha}.$$

Let us now bound the bias term B_{II}. We have

$$B_{II}(t) = \bar{B}_{II}(t) + \tilde{B}_{II}(t),$$

where

$$\bar{B}_{II}(t) = \sum_{i=1}^{n} h_n^{-1}\delta_i(t) \text{ and } \tilde{B}_{II}(t) = \sum_{i=1}^{n} h_n^{-1} K\Big(\frac{t-t_{i-1}}{h_n}\Big)\gamma_i(t),$$

with

$$\delta_i(t) = \mathbb{E}\Big(K\Big(\frac{t-t_{i-1}}{h_n}\Big)\int_{t_{i-1}}^{t_i} e^{-2\vartheta(t-u)}\sigma_u^2 du\Big|\mathcal{G}_{i-1}\Big) - \int_{t_{i-1}}^{t_i} K\Big(\frac{t-u}{h_n}\Big)e^{-2\vartheta(t-u)}\sigma_u^2 du,$$

and

$$\gamma_i(t) = \mathbb{E}\Big(\frac{\big((\Delta_i^n Y^1)^2 - (\Delta_i^n Y^2)^2\big)}{e^{-2\vartheta(T_1-t)} - e^{-2\vartheta(T_2-t)}} - \int_{t_{i-1}}^{t_i} e^{-2\vartheta(t-u)}\sigma_u^2 du\Big|\mathcal{G}_{i-1}\Big).$$

In Section 2.4.3, we got

$$\sup_{t\in[h_n,T]} \mathbb{E}\big((\bar{B}_{II}(t))^2\big) \lesssim \Delta_n h_n^{-1}.$$

The term $\tilde{B}_{II}(t)$ cannot be treated in the same way as in that section, as the expectation does not cancel. We have

$$\gamma_i(t) = \frac{1}{e^{-2\vartheta(T_1-t)} - e^{-2\vartheta(T_2-t)}}\mathbb{E}(A_i^n + B_i^n + C_i^n + D_i^n + E_i^n + F_i^n|\mathcal{G}_{i-1}),$$

where

$$\begin{aligned}
A_i^n &= \Big(\int_{t_{i-1}}^{t_i} b_t^1 dt\Big)^2 - \Big(\int_{t_{i-1}}^{t_i} b_t^2 dt\Big)^2,\\
B_i^n &= 2\int_{t_{i-1}}^{t_i} b_t^1 dt\Big(\int_{t_{i-1}}^{t_i} e^{-\vartheta(T_1-t)}\sigma_t dB_t + \int_{t_{i-1}}^{t_i} \overline{\sigma}_t d\overline{B}_t + \kappa_1^n \Delta_i^n \epsilon^1\Big),\\
C_i^n &= -2\int_{t_{i-1}}^{t_i} b_t^2 dt\Big(\int_{t_{i-1}}^{t_i} e^{-\vartheta(T_2-t)}\sigma_t dB_t + \int_{t_{i-1}}^{t_i} \overline{\sigma}_t d\overline{B}_t + \kappa_2^n \Delta_i^n \epsilon^2\Big),\\
D_i^n &= \big(\kappa_1^n \Delta_i^n \epsilon^1 - \kappa_2^n \Delta_i^n \epsilon^2\big)^2,\\
E_i^n &= 2\kappa_1^n \Delta_i^n \epsilon^1\Big(\int_{t_{i-1}}^{t_i} e^{-\vartheta(T_1-t)}\sigma_t dB_t + \int_{t_{i-1}}^{t_i} \overline{\sigma}_t d\overline{B}_t\Big),\\
F_i^n &= -2\kappa_2^n \Delta_i^n \epsilon^2\Big(\int_{t_{i-1}}^{t_i} e^{-\vartheta(T_2-t)}\sigma_t dB_t + \int_{t_{i-1}}^{t_i} \overline{\sigma}_t d\overline{B}_t\Big).
\end{aligned}$$

Because, by localization, the processes b, σ and $\overline{\sigma}$ can be assumed to be bounded, and as $\beta > 1/2$, the terms $\mathbb{E}\big((A_i^n)^2\big)$ and $\mathbb{E}\big((B_i^n)^2 + (C_i^n)^2\big)$ have respective orders Δ_n^4 and Δ_n^3. Besides, $\mathbb{E}\big((D_i^n)^2\big)$ has order $\Delta_n^{4\beta}$, and

$$\begin{aligned}\mathbb{E}(E_i^n|\mathcal{G}_{i-1}) =&\mathbb{E}(\mathbb{E}(E_i^n|\mathcal{H}_{i-1})|\mathcal{G}_{i-1})\\ =&-2\kappa_1^n\epsilon_{i-1}^1\mathbb{E}\Big(\int_{t_{i-1}}^{t_i} e^{-\vartheta(T_1-t)}\sigma_t dB_t + \int_{t_{i-1}}^{t_i}\overline{\sigma}_t d\overline{B}_t\Big|\mathcal{G}_{i-1}\Big)\\ =&0,\end{aligned}$$

and $\mathbb{E}(F_i^n|\mathcal{G}_{i-1}) = 0$ too. Now, $\mathbb{E}(\gamma_i(t)^2)$ is less than

$$\begin{aligned}&4\Big(\frac{1}{e^{-2\vartheta(T_1-t)} - e^{-2\vartheta(T_2-t)}}\Big)^2\mathbb{E}\big(\mathbb{E}(A_i^n|\mathcal{G}_{i-1})^2 + \mathbb{E}(B_i^n|\mathcal{G}_{i-1})^2 + \mathbb{E}(C_i^n|\mathcal{G}_{i-1})^2 + \mathbb{E}(D_i^n|\mathcal{G}_{i-1})^2\big)\\ \leq&4\Big(\frac{1}{e^{-2\vartheta(T_1-t)} - e^{-2\vartheta(T_2-t)}}\Big)^2\big(\mathbb{E}\big((A_i^n)^2\big) + \mathbb{E}\big((B_i^n)^2\big) + \mathbb{E}\big((C_i^n)^2\big) + \mathbb{E}\big((D_i^n)^2\big)\big)\end{aligned}$$

by Jensen inequality, so that $\mathbb{E}(\gamma_i(t)^2)$ is of order Δ_n^3 if $\beta \geq 3/4$, and of order $\Delta_n^{4\beta}$ if $1/2 < \beta < 3/4$. It follows that

$$\begin{aligned}\mathbb{E}\big((\tilde{B}_{II}(t))^2\big) =&\sum_{i=1}^n h_n^{-2}K\Big(\frac{t-t_{i-1}}{h_n}\Big)^2\mathbb{E}(\gamma_i(t)^2)\\ &+2\sum_{1\leq i<j\leq n} h_n^{-2}K\Big(\frac{t-t_{i-1}}{h_n}\Big)K\Big(\frac{t-t_{j-1}}{h_n}\Big)\mathbb{E}(\gamma_i(t)\gamma_j(t))\\ \leq&\sum_{i=1}^n h_n^{-2}K\Big(\frac{t-t_{i-1}}{h_n}\Big)^2\mathbb{E}(\gamma_i(t)^2)\\ &+2\sum_{1\leq i<j\leq n} h_n^{-2}K\Big(\frac{t-t_{i-1}}{h_n}\Big)K\Big(\frac{t-t_{j-1}}{h_n}\Big)\sqrt{\mathbb{E}(\gamma_i(t)^2)\mathbb{E}(\gamma_j(t)^2)}\end{aligned}$$

by Tchebychev inequality. There are at most $O(h_n^{-1}\Delta_n)$ terms that are not zero in the first sum in the RHS, because K has compact support; there are at most $O(h_n^{-2}\Delta_n^2)$ non-zero terms in the other sum in the RHS. Therefore, we have bounded $\mathbb{E}\big((\tilde{B}_{II}(t))^2\big)$ by a term that is of order Δ_n if $\beta \geq 3/4$, and of order $\Delta_n^{4\beta-2}$ if $1/2 < \beta < 3/4$. We want to know if that last bound is the lowest possible: let us consider a very simple submodel with $b^1 = b^2 = 0$, $\sigma = 0$, $\kappa_1^n = \kappa_2^n = \Delta_n^\beta$ and ϵ_i^1, ϵ_i^2 being independent random variables with law $\mathcal{N}(0,1)$. Then

$$\gamma_i(t) = \Delta_n^{2\beta}\frac{(\epsilon_{i-1}^1)^2 - (\epsilon_{i-1}^2)^2}{e^{-2\vartheta(T_1-t)} - e^{-2\vartheta(T_2-t)}},$$

so that within this submodel, $\mathbb{E}\big((\tilde{B}_{II}(t))^2\big)$ is exactly of order $\Delta_n^{4\beta-2}$. This is therefore the sharpest bound for $\mathbb{E}\big((\tilde{B}_{II}(t))^2\big)$ in the general model. We may also see that this bound is uniform on $t \in [h_n, T]$.

Let us summarize: we got $\sup_{t\in[h_n,T]}\mathbb{E}\big[(V_n(t))^2\big]\lesssim \Delta_n h_n^{-1}$, $\sup_{t\in[h_n,T]}\mathbb{E}\big((B_I(t))^2\big)\lesssim h_n^{2\alpha}$ and finally $\sup_{t\in[h_n,T]}\mathbb{E}\big((B_{II}(t))^2\big)=O\big(\Delta_n h_n^{-1}\vee\Delta_n^{1\wedge(4\beta-2)}\big)$. Recall that the term I was of order $\Delta_n^{1/2\wedge(2\beta-1)}$; it is thus either possible to reach the optimal rate $\Delta_n^{-2\alpha/(2\alpha+1)}$ with the choice $h_n=\Delta_n^{1/(2\alpha+1)}$, or impossible because $\Delta_n^{4\beta-2}$ is predominant over $\Delta_n^{2\alpha/(2\alpha+1)}$. We may choose $h_n=\Delta_n^{1/(2\alpha+1)}$ anyway, which allows to get the optimal rate whenever possible.
We conclude the proof in the same way for $\overline{\sigma}$.

3.4.6 Proof of Theorem 3.5

Let us look at the properties of the estimator $\widetilde{\sigma}^2_{3,n,t}$. For ease of notation, we write $\bar{\vartheta}_{3,n}$ for $\max\{\vartheta_{3,n},\varpi_n\}$. We have

$$\widetilde{\sigma}^2_{3,n,t}-\sigma_t^2=\frac{\sum_{i=1}^n K_{h_n}\big(t-t_{i-1}\big)\big(\Delta_i^nY^1\Delta_i^nY^2-\Delta_i^nY^2\Delta_i^nY^3\big)}{e^{-\bar{\vartheta}_{3,n}(T_1+T_2-2t)}-e^{-\bar{\vartheta}_{3,n}(T_1+T_3-2t)}}-\sigma_t^2=I+II,$$

with

$$\begin{aligned}I=&\Big(\frac{1}{e^{-\bar{\vartheta}_{3,n}(T_1+T_2-2t)}-e^{-\bar{\vartheta}_{3,n}(T_1+T_3-2t)}}-\frac{1}{e^{-\vartheta(T_1+T_2-2t)}-e^{-\vartheta(T_1+T_3-2t)}}\Big)\\&\times\sum_{i=1}^n K_{h_n}\big(t-t_{i-1}\big)\big(\Delta_i^nY^1\Delta_i^nY^2-\Delta_i^nY^1\Delta_i^nY^3\big)\end{aligned}$$

and

$$II=\frac{\sum_{i=1}^n K_{h_n}\big(t-t_{i-1}\big)\big(\Delta_i^nY^1\Delta_i^nY^2-\Delta_i^nY^1\Delta_i^nY^3\big)}{e^{-\vartheta(T_1+T_2-2t)}-e^{-\vartheta(T_1+T_3-2t)}}-\sigma_t^2.$$

The term I Since, for $1\le j,k\le 3, j\neq k$,

$$\Delta_i^nY^j\Delta_i^nY^k=\Delta_i^nX^j\Delta_iX^k+\kappa_j^n\Delta_i^n\epsilon^j\Delta_i^nX^k+\kappa_k^n\Delta_i^n\epsilon^k\Delta_i^nX^j+\kappa_j^n\Delta_i^n\epsilon^j\kappa_k^n\Delta_i^n\epsilon^k,$$

and because $\mathbb{E}(|\Delta_i^nX^j\Delta_iX^k|)$, $\mathbb{E}(|\kappa_j^n\Delta_i^n\epsilon^j\Delta_i^nX^k|)$ and $E(|\kappa_j^n\Delta_i^n\epsilon^j\kappa_k^n\Delta_i^n\epsilon^k|)$ have respective orders Δ_n, $\Delta_n^{\beta+1/2}$ and $\Delta_n^{2\beta}$, $\mathbb{E}(|\Delta_i^nY^j\Delta_i^nY^k|)$ has order Δ_n and

$$\mathbb{E}\Big[\Big|\sum_{i=1}^n K_{h_n}(t-t_{i-1})\big(\Delta_i^nY^1\Delta_i^nY^2-\Delta_i^nY^1\Delta_i^nY^3\big)\Big|\Big]\lesssim\sum_{i=1}^n K_{h_n}(t-t_{i-1})\Delta_n\lesssim 1$$

since $K_{h_n}(t-t_{i-1})$ is of order h_n^{-1} for a number of terms that are at most of order $\Delta_n^{-1}h_n$. Therefore $\sum_{i=1}^n K_{h_n}\big(t-t_{i-1}\big)\big(\Delta_i^nY^1\Delta_i^nY^2-\Delta_i^nY^1\Delta_i^nY^3\big)$ is bounded in expectation hence tight. The order of I is thus dictated by

$$\frac{1}{e^{-2\bar{\vartheta}_{3,n}(T_1-t)}-e^{-2\bar{\vartheta}_{3,n}(T_2-t)}}-\frac{1}{e^{-2\vartheta(T_1-t)}-e^{-2\vartheta(T_2-t)}}$$

which, by Theorem 3.3, is of order $\Delta_n^{1/2}$.

The term II The term II further splits into $II = (e^{-\vartheta(T_1+T_2-2t)} - e^{-\vartheta(T_1+T_3-2t)})^{-1}\big(B_n(t) + V_n(t)\big)$, where $V_n(t)$ is

$$\sum_{i=1}^{n} K_{h_n}(t-t_{i-1})\big(\Delta_i^n Y^1 \Delta_i^n Y^2 - \Delta_i^n Y^1 \Delta_i^n Y^3 - \mathbb{E}\big[\Delta_i^n Y^1 \Delta_i^n Y^2 - \Delta_i^n Y^1 \Delta_i^n Y^3 \big| \mathcal{G}_{i-1}\big]\big)$$

and $B_n(t)$ is

$$\sum_{i=1}^{n} \mathbb{E}\big[K_{h_n}(t-t_{i-1})\big(\Delta_i^n Y^1 \Delta_i^n Y^2 - \Delta_i^n Y^1 \Delta_i^n Y^3\big)\big|\mathcal{G}_{i-1}\big] - \big(e^{-\vartheta(T_1+T_2-2t)} - e^{-\vartheta(T_1+T_3-2t)}\big)\sigma_t^2.$$

Bounding the variance term

We first prove an upper bound for $\mathbb{E}(V_n(t)^2)$ uniformly in $t \in [h_n, T]$. When examining the term I, we said that for $1 \leq j, k \leq 3, j \neq k$,

$$\Delta_i^n Y^j \Delta_i^n Y^k = \Delta_i^n X^j \Delta_i X^k + \kappa_j^n \Delta_i^n \epsilon^j \Delta_i^n X^k + \kappa_k^n \Delta_i^n \epsilon^k \Delta_i^n X^j + \kappa_j^n \Delta_i^n \epsilon^j \kappa_k^n \Delta_i^n \epsilon^k,$$

and the order of $\mathbb{E}((\Delta_i^n Y^j \Delta_i^n Y^k)^2)$ is the one of $\mathbb{E}((\Delta_i^n X^j \Delta_i^n X^k)^2)$, that is Δ_n^2. It follows that

$$\begin{aligned}
&\sup_{t\in[h_n,T]} \mathbb{E}\big((V_n(t))^2\big)\\
&= \sup_{t\in[h_n,T]} h_n^{-2} \sum_{i=1}^{n} K^2\Big(\frac{t-t_{i-1}}{h_n}\Big)\mathbb{E}\Big[\big[\Delta_i^n Y^1 \Delta_i^n Y^2 - \Delta_i^n Y^1 \Delta_i^n Y^3\\
&\qquad - \mathbb{E}\big(\Delta_i^n Y^1 \Delta_i^n Y^2 - \Delta_i^n Y^1 \Delta_i^n Y^3 \big| \mathcal{G}_{i-1}\big)\big]^2\Big]
\end{aligned}$$

is of order $\Delta_n h_n^{-1}$, as K has compact support.

Bounding the bias term

In order to get an upper bound for the bias term $\sup_{t\in[h_n,T]} \mathbb{E}\big((B_n(t))^2\big)$, we use the decomposition

$$B_n(t) = \big(e^{-\vartheta(T_1+T_2-2t)} - e^{-\vartheta(T_1+T_3-2t)}\big)(B_I(t) + B_{II}(t)),$$

where

$$B_I(t) = \int_0^T h_n^{-1} K\Big(\frac{t-u}{h_n}\Big) e^{-2\vartheta(t-u)}\sigma_u^2 du - \sigma_t^2$$

and

$$\begin{aligned}
B_{II}(t) =& \frac{\sum_{i=1}^{n} \mathbb{E}\Big(h_n^{-1} K\Big(\frac{t-t_{i-1}}{h_n}\Big)\big(\Delta_i^n Y^1 \Delta_i^n Y^2 - \Delta_i^n Y^1 \Delta_i^n Y^3\big)\Big|\mathcal{G}_{i-1}\Big)}{e^{-\vartheta(T_1+T_2-2t)} - e^{-\vartheta(T_1+T_3-2t)}}\\
&- \int_0^T h_n^{-1} K\Big(\frac{t-u}{h_n}\Big) e^{-2\vartheta(t-u)}\sigma_u^2 du.
\end{aligned}$$

The term B_I has already been treated:

$$\sup_{t\in[h_n,T]} \mathbb{E}\big((B_I(t))^2\big) \lesssim h_n^{2\alpha}.$$

Let us now bound the bias term B_{II}. We have

$$B_{II}(t) = \bar{B}_{II}(t) + \tilde{B}_{II}(t),$$

where

$$\bar{B}_{II}(t) = \sum_{i=1}^{n} h_n^{-1}\delta_i(t) \text{ and } \tilde{B}_{II}(t) = \sum_{i=1}^{n} h_n^{-1} K\Big(\frac{t-t_{i-1}}{h_n}\Big)\gamma_i(t),$$

with

$$\delta_i(t) = \mathbb{E}\Big(K\Big(\frac{t-t_{i-1}}{h_n}\Big)\int_{t_{i-1}}^{t_i} e^{-2\vartheta(t-u)}\sigma_u^2 du\Big|\mathcal{G}_{i-1}\Big) - \int_{t_{i-1}}^{t_i} K\Big(\frac{t-u}{h_n}\Big)e^{-2\vartheta(t-u)}\sigma_u^2 du,$$

and

$$\gamma_i(t) = \mathbb{E}\Big(\frac{\big(\Delta_i^n Y^1\Delta_i^n Y^2 - \Delta_i^n Y^1 \Delta_i^n Y^3\big)}{e^{-\vartheta(T_1+T_2-2t)} - e^{-\vartheta(T_1+T_3-2t)}} - \int_{t_{i-1}}^{t_i} e^{-2\vartheta(t-u)}\sigma_u^2 du\Big|\mathcal{G}_{i-1}\Big).$$

We have already proved that

$$\sup_{t\in[h_n,T]} \mathbb{E}\big((\bar{B}_{II}(t))^2\big) \lesssim \Delta_n h_n^{-1}.$$

Now we care for the term $\tilde{B}_{II}(t)$. We have

$$\gamma_i(t) = \frac{1}{e^{-\vartheta(T_1+T_2-2t)} - e^{-\vartheta(T_1+T_3-2t)}}\mathbb{E}(A_i^n + B_i^n + C_i^n + D_i^n + E_i^n|\mathcal{G}_{i-1}),$$

where

$$\begin{aligned}
A_i^n &= \int_{t_{i-1}}^{t_i} b_t^1 dt \int_{t_{i-1}}^{t_i} (b_t^2 - b_t^3)dt,\\
B_i^n &= \mathfrak{e}_{3,2}(\vartheta)\int_{t_{i-1}}^{t_i} b_t^1 dt \int_{t_{i-1}}^{t_i} e^{\vartheta t}\sigma_t dB_t,\\
C_i^n &= \int_{t_{i-1}}^{t_i} (b_t^2 - b_t^3)dt\Big(\int_{t_{i-1}}^{t_i} e^{-\vartheta(T_1-t)}\sigma_t dB_t + \int_{t_{i-1}}^{t_i} \overline{\sigma}_t d\overline{B}_t\Big),\\
D_i^n &= (\kappa_3^n\epsilon_{i-1}^3 - \kappa_2^n\epsilon_{i-1}^2)\int_{t_{i-1}}^{t_i} b_t^1 dt + \kappa_1^n\epsilon_{i-1}^1\int_{t_{i-1}}^{t_i}(b_t^3 - b_t^2)dt,\\
E_i^n &= \kappa_1^n\kappa_2^n\epsilon_{i-1}^1\epsilon_{i-1}^2 - \kappa_1^n\kappa_3^n\epsilon_{i-1}^1\epsilon_{i-1}^3.
\end{aligned}$$

Because, by localization, the processes b, σ and $\overline{\sigma}$ can be assumed to be bounded, and as $\beta \geq 1/2$, the terms $\mathbb{E}\big((A_i^n))^2\big)$ and $\mathbb{E}\big((B_i^n)^2 + (C_i^n)^2\big)$ have respective orders Δ_n^4 and Δ_n^3.

Besides, $\mathbb{E}\big((D_i^n)^2\big)$ has order $\Delta_n^{2+2\beta}$ and $\mathbb{E}\big((E_i^n)^2\big)$ has order $\Delta_n^{4\beta}$. Notice that $\mathbb{E}\big(E_i^n E_j^n\big) = 0$ when $i \neq j$.
As $\beta \geq 1/2$, it appears that $\mathbb{E}((A_i^n)^2 + (B_i^n)^2 + (C_i^n)^2 + (D_i^n)^2)$ is of order Δ_n^3. By convexity inequality,

$$
\begin{aligned}
\mathbb{E}\big((\tilde{B}_{II}(t))^2\big) =& \mathbb{E}\Big(\Big(\sum_{i=1}^n h_n^{-1} K\Big(\frac{t-t_{i-1}}{h_n}\Big)\mathbb{E}(A_i^n + B_i^n + C_i^n + D_i^n + E_i^n|\mathcal{G}_{i-1})\Big)^2\Big) \\
\leq& 2\mathbb{E}\Big(\Big(\sum_{i=1}^n h_n^{-1} K\Big(\frac{t-t_{i-1}}{h_n}\Big)\mathbb{E}(A_i^n + B_i^n + C_i^n + D_i^n|\mathcal{G}_{i-1})\Big)^2\Big) \\
&+ 2\mathbb{E}\Big(\Big(\sum_{i=1}^n h_n^{-1} K\Big(\frac{t-t_{i-1}}{h_n}\Big)\mathbb{E}(E_i^n|\mathcal{G}_{i-1})\Big)^2\Big).
\end{aligned}
$$

We have

$$
\begin{aligned}
&\mathbb{E}\Big(\Big(\sum_{i=1}^n h_n^{-1} K\Big(\frac{t-t_{i-1}}{h_n}\Big)\mathbb{E}(A_i^n + B_i^n + C_i^n + D_i^n|\mathcal{G}_{i-1})\Big)^2\Big) \\
=& \sum_{i=1}^n h_n^{-2} K\Big(\frac{t-t_{i-1}}{h_n}\Big)^2 \mathbb{E}\big(\mathbb{E}(A_i^n + B_i^n + C_i^n + D_i^n|\mathcal{G}_{i-1})^2\big) \\
&+ 2 \sum_{1\leq i<j\leq n} h_n^{-2} K\Big(\frac{t-t_{i-1}}{h_n}\Big) K\Big(\frac{t-t_{j-1}}{h_n}\Big) \\
&\qquad \times \mathbb{E}\big(\mathbb{E}(A_i^n + B_i^n + C_i^n + D_i^n|\mathcal{G}_{i-1})\mathbb{E}(A_j^n + B_j^n + C_j^n + D_j^n|\mathcal{G}_{j-1})\big) \\
\leq& \sum_{i=1}^n h_n^{-2} K\Big(\frac{t-t_{i-1}}{h_n}\Big)^2 4\mathbb{E}\big((A_i^n)^2 + (B_i^n)^2 + (C_i^n)^2 + (D_i^n)^2\big) \\
&+ 2 \sum_{1\leq i<j\leq n} h_n^{-2} K\Big(\frac{t-t_{i-1}}{h_n}\Big) K\Big(\frac{t-t_{j-1}}{h_n}\Big) \\
&\qquad \times \sqrt{16\mathbb{E}\big((A_i^n)^2 + (B_i^n)^2 + (C_i^n)^2 + (D_i^n)^2\big)\mathbb{E}\big((A_j^n)^2 + (B_j^n)^2 + (C_j^n)^2 + (D_j^n)^2\big)}
\end{aligned}
$$

by Tchebychev inequality. The whole term is therefore of order Δ_n, and

$$
\begin{aligned}
&\mathbb{E}\Big(\Big(\sum_{i=1}^n h_n^{-1} K\Big(\frac{t-t_{i-1}}{h_n}\Big)\mathbb{E}(E_i^n|\mathcal{G}_{i-1})\Big)^2\Big) \\
=& \mathbb{E}\Big(\Big(\sum_{i=1}^n h_n^{-1} K\Big(\frac{t-t_{i-1}}{h_n}\Big)\mathbb{E}(E_i^n)\Big)^2\Big) \\
=& \sum_{i=1}^n h_n^{-2} K\Big(\frac{t-t_{i-1}}{h_n}\Big)^2 \mathbb{E}\big((E_i^n)^2\big) + 2 \sum_{1\leq i<j\leq n} h_n^{-2} K\Big(\frac{t-t_{i-1}}{h_n}\Big) K\Big(\frac{t-t_{j-1}}{h_n}\Big)\mathbb{E}\big(E_i^n E_j^n\big).
\end{aligned}
$$

We noted that $\mathbb{E}\big(E_i^n E_j^n\big) = 0$, and as $\mathbb{E}\big((E_i^n)^2\big)$ has order $\Delta_n^{4\beta}$ and K has compact support, there remains

$$\mathbb{E}\Big(\Big(\sum_{i=1}^n h_n^{-1} K\Big(\frac{t-t_{i-1}}{h_n}\Big)\mathbb{E}(E_i^n|\mathcal{G}_{i-1})\Big)^2\Big) \lesssim n h_n^{-1}\Delta_n^{4\beta} \simeq \Delta_n^{4\beta-1} h_n^{-1} = \Delta_n h_n^{-1}\Delta_n^{4\beta-2}.$$

Because $\beta \geq 1/2$, $\Delta_n^{4\beta-2} = O(1)$ as $n \to \infty$, so that

$$\mathbb{E}\Big(\Big(\sum_{i=1}^n h_n^{-1} K\Big(\frac{t-t_{i-1}}{h_n}\Big)\mathbb{E}(E_i^n|\mathcal{G}_{i-1})\Big)^2\Big) \lesssim \Delta_n h_n^{-1},$$

and finally,

$$\sup_{t\in[h_n,T]} \mathbb{E}\big((B_{II}(t))^2\big) \leq 2 \sup_{t\in[h_n,T]} \mathbb{E}\big((\bar{B}_{II}(t))^2\big) + 2 \sup_{t\in[h_n,T]} \mathbb{E}\big((\tilde{B}_{II}(t))^2\big) \lesssim \Delta_n h_n^{-1}.$$

As we have got $\sup_{t\in[h_n,T]} \mathbb{E}\big[(V_n(t))^2\big] \lesssim \Delta_n h_n^{-1}$, as well as $\sup_{t\in[h_n,T]} \mathbb{E}\big((B_I(t))^2\big) \lesssim h_n^{2\alpha}$ and $\sup_{t\in[h_n,T]} \mathbb{E}\big((B_{II}(t))^2\big) \lesssim \Delta_n h_n^{-1}$, all error terms are of order $\Delta_n^{2\alpha/(2\alpha+1)}$ with the choice $h_n = \Delta_n^{1/(2\alpha+1)}$.
We conclude the proof in the same way for $\overline{\sigma}$.

3.4.7 Proof of Theorem 3.6

Let $\widetilde{\ell}_{\vartheta,\sigma,\overline{\sigma}}$ denote

$$\frac{(\Delta_i^n Y^2 - \Delta_i^n Y^1)(\Delta_i^n Y^2 - e^{-\vartheta(T_2-T_1)}\Delta_i^n Y^1)e^{-\vartheta(T_2-T_1)}(T_2-T_1)}{(1-e^{-\vartheta(T_2-T_1)})^3 \int_{(i-1)\Delta_n}^{i\Delta_n} \overline{\sigma}_t^2 dt},$$

which is not an efficient score function anymore when model errors are included, but is the analog of the efficient score $\widetilde{\ell}_{\vartheta,\sigma,\overline{\sigma}}$ in Chapter 2.

First step

Let ϑ_n be a deterministic sequence such that $\sqrt{n}(\vartheta_n - \vartheta) = O(1)$. First we show that $\Delta_n^{1/2}\sum_{i\in\mathcal{I}_n}\big(\widetilde{\ell}^i_{\vartheta_n,\sigma,\overline{\sigma}} - \widetilde{\ell}\big(\vartheta_n, \widehat{\overline{\sigma}}_n^2\big)^i\big) \to 0$ in probability, as $n\to\infty$. This amounts to show that

$$\Delta_n^{1/2}\sum_{i\in\mathcal{I}_n}(\Delta_i^n Y^2 - \Delta_i^n Y^1)(\Delta_i^n Y^2 - e^{-\vartheta_n(T_2-T_1)}\Delta_i^n Y^1)\Big(\frac{1}{\int_{t_{i-1}}^{t_i}\overline{\sigma}_t^2 dt} - \frac{1}{\Delta_n \widehat{\overline{\sigma}}^2_{n,t_{i-1}}}\Big)$$

converges to 0 in probability. We rewrite it as

$$\begin{aligned}
&\Delta_n^{1/2}\sum_{i\in\mathcal{I}_n}(\Delta_i^n Y^2 - \Delta_i^n Y^1)(\Delta_i^n Y^2 - e^{-\vartheta_n(T_2-T_1)}\Delta_i^n Y^1)\Big(\frac{1}{\int_{t_{i-1}}^{t_i}\overline{\sigma}_t^2 dt} - \frac{1}{\Delta_n \overline{\sigma}^2_{t_{i-1}}}\Big)\\
&+ \Delta_n^{1/2}\sum_{i\in\mathcal{I}_n}(\Delta_i^n Y^2 - \Delta_i^n Y^1)(\Delta_i^n Y^2 - e^{-\vartheta_n(T_2-T_1)}\Delta_i^n Y^1)\frac{1}{\Delta_n}\Big(\frac{1}{\overline{\sigma}^2_{t_{i-1}}} - \frac{1}{\widehat{\overline{\sigma}}^2_{n,t_{i-1}}}\Big)\\
=&S_n' + S_n''
\end{aligned}$$

where

$$S'_n = \Delta_n^{-1/2} \sum_{i \in \mathcal{I}_n} (\Delta_i^n Y^2 - \Delta_i^n Y^1)(\Delta_i^n Y^2 - e^{-\vartheta_n (T_2 - T_1)} \Delta_i^n Y^1) \frac{\Delta_n \overline{\sigma}^2_{t_{i-1}} - \int_{t_{i-1}}^{t_i} \overline{\sigma}^2_t dt}{\overline{\sigma}^2_{t_{i-1}} \int_{t_{i-1}}^{t_i} \overline{\sigma}^2_t dt}$$

and

$$S''_n = \Delta_n^{-1/2} \sum_{i \in \mathcal{I}_n} (\Delta_i^n Y^2 - \Delta_i^n Y^1)(\Delta_i^n Y^2 - e^{-\vartheta_n (T_2 - T_1)} \Delta_i^n Y^1) \frac{\widehat{\overline{\sigma}}^2_{t_{i-1}} - \overline{\sigma}^2_{t_{i-1}}}{\overline{\sigma}^2_{t_{i-1}} \widehat{\overline{\sigma}}^2_{t_{i-1}}}.$$

To care for S'_n, we have that

$$\mathbb{E}(|S'_n|) \leq \sum_{i \in \mathcal{I}_n} \mathbb{E}\Big(\Big| \Delta_n^{-1/2} (\Delta_i^n Y^2 - \Delta_i^n Y^1)(\Delta_i^n Y^2 - e^{-\vartheta_n (T_2 - T_1)} \Delta_i^n Y^1) \frac{\Delta_n \overline{\sigma}^2_{t_{i-1}} - \int_{t_{i-1}}^{t_i} \overline{\sigma}^2_t dt}{\overline{\sigma}^2_{t_{i-1}} \int_{t_{i-1}}^{t_i} \overline{\sigma}^2_t dt} \Big| \Big),$$

and for $i \in \mathcal{I}_n$,

$$\begin{aligned} &\mathbb{E}\Big(\Big| \Delta_n^{-1/2} (\Delta_i^n Y^2 - \Delta_i^n Y^1)(\Delta_i^n Y^2 - e^{-\vartheta_n (T_2 - T_1)} \Delta_i^n Y^1) \frac{\Delta_n \overline{\sigma}^2_{t_{i-1}} - \int_{t_{i-1}}^{t_i} \overline{\sigma}^2_t dt}{\overline{\sigma}^2_{t_{i-1}} \int_{t_{i-1}}^{t_i} \overline{\sigma}^2_t dt} \Big| \Big) \\ \leq & \frac{\Delta_n^{-3/2}}{\tilde{c}^4} \sqrt{\mathbb{E}\big(\big| (\Delta_i^n Y^2 - \Delta_i^n Y^1)(\Delta_i^n Y^2 - e^{-\vartheta_n (T_2 - T_1)} \Delta_i^n Y^1) \big|^2 \big) \mathbb{E}\Big(\Big| \Delta_n \overline{\sigma}^2_{t_{i-1}} - \int_{t_{i-1}}^{t_i} \overline{\sigma}^2_t dt \Big|^2 \Big)} \end{aligned}$$

using Cauchy-Schwarz and the fact that $\mathbb{P}((\sigma, \overline{\sigma}) \in \Sigma(c, \tilde{c})) = 1$. We have that $\mathbb{E}\big(\big| (\Delta_i^n Y^2 - \Delta_i^n Y^1)(\Delta_i^n Y^2 - e^{-\vartheta_n (T_2 - T_1)} \Delta_i^n Y^1) \big|^2 \big)$ is of order Δ_n^2, because $\mathbb{E}\big(\big| (\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta_n (T_2 - T_1)} \Delta_i^n X^1) \big|^2 \big)$, which is the predominant term in that expectation, is itself of order Δ_n^2, using BDG inequality. Also,

$$\begin{aligned} &\mathbb{E}\Big(\Big| \Delta_n^{-1/2} (\Delta_i^n Y^2 - \Delta_i^n Y^1)(\Delta_i^n Y^2 - e^{-\vartheta_n (T_2 - T_1)} \Delta_i^n Y^1) \frac{\Delta_n \overline{\sigma}^2_{t_{i-1}} - \int_{t_{i-1}}^{t_i} \overline{\sigma}^2_t dt}{\overline{\sigma}^2_{t_{i-1}} \int_{t_{i-1}}^{t_i} \overline{\sigma}^2_t dt} \Big| \Big) \\ \lesssim & \Delta_n^{-1/2} \sqrt{\mathbb{E}\Big(\Big| \Delta_n \overline{\sigma}^2_{t_{i-1}} - \int_{t_{i-1}}^{t_i} \overline{\sigma}^2_t dt \Big|^2 \Big)} \\ \lesssim & \Delta_n^{\alpha + 1/2}. \end{aligned}$$

Finally, $\mathbb{E}(|S'_n|) \lesssim \Delta_n^{\alpha - 1/2}$, so that S'_n converges to 0 in L^1 and thus in probability.

Now we look at the term S''_n : because the kernel used for nonparametric estimation has its support included in $(0,+\infty)$, each $\widehat{\overline{\sigma}}^2_{n,t_{i-1}}$ is $\mathcal{G}_{i-1}$-measurable, and

$$\begin{aligned}
&\mathbb{E}\Big(\Delta_n^{-1/2}(\Delta_i^n Y^2-\Delta_i^n Y^1)(\Delta_i^n Y^2-e^{-\vartheta_n(T_2-T_1)}\Delta_i^n Y^1)\frac{\widehat{\overline{\sigma}}^2_{n,t_{i-1}}-\overline{\sigma}^2_{t_{i-1}}}{\overline{\sigma}^2_{t_{i-1}}\widehat{\overline{\sigma}}^2_{n,t_{i-1}}}\Big|\mathcal{G}_{i-1}\Big)\\
=&\Delta_n^{-1/2}\mathbb{E}\big((\Delta_i^n Y^2-\Delta_i^n Y^1)(\Delta_i^n Y^2-e^{-\vartheta_n(T_2-T_1)}\Delta_i^n Y^1)\big|\mathcal{G}_{i-1}\big)\frac{\widehat{\overline{\sigma}}^2_{n,t_{i-1}}-\overline{\sigma}^2_{t_{i-1}}}{\overline{\sigma}^2_{t_{i-1}}\widehat{\overline{\sigma}}^2_{n,t_{i-1}}}\\
=&\Delta_n^{-1/2}\Big(\big(e^{-\vartheta(T_2-T_1)}-e^{-\vartheta_n(T_2-T_1)}\big)\chi_i^n+A_i^n+B_i^n+C_i^n+D_i^n\Big)\frac{\widehat{\overline{\sigma}}^2_{n,t_{i-1}}-\overline{\sigma}^2_{t_{i-1}}}{\overline{\sigma}^2_{t_{i-1}}\widehat{\overline{\sigma}}^2_{n,t_{i-1}}},
\end{aligned}$$

where

$$\begin{aligned}
\chi_i^n=&\big(e^{-\vartheta(T_2-T_1)}-1\big)\mathbb{E}\Big(\int_{t_{i-1}}^{t_i}e^{-2\vartheta(T_1-t)}\sigma_t^2dt\Big|\mathcal{G}_{i-1}\Big),\\
A_i^n=&\int_{t_{i-1}}^{t_i}(b_t^2-b_t^1)dt\int_{t_{i-1}}^{t_i}\big(b_t^2-e^{-\vartheta_n(T_2-T_1)}b_t^1\big)dt,\\
B_i^n=&\int_{t_{i-1}}^{t_i}(b_t^2-b_t^1)dt\Big(\int_{t_{i-1}}^{t_i}\big(e^{-\vartheta(T_2-T_1)}-e^{-\vartheta_n(T_2-T_1)}\big)e^{-\vartheta(T_1-t)}\sigma_t dB_t\\
&\qquad+\int_{t_{i-1}}^{t_i}\big(1-e^{-\vartheta_n(T_2-T_1)}\big)\overline{\sigma}_t d\overline{B}_t+\kappa_2^n\Delta_i^n\epsilon^2-e^{-\vartheta_n(T_2-T_1)}\kappa_1^n\Delta_i^n\epsilon^1\Big),\\
C_i^n=&\int_{t_{i-1}}^{t_i}(b_t^2-e^{-\vartheta_n(T_2-T_1)}b_t^1)dt\Big(\int_{t_{i-1}}^{t_i}\big(e^{-\vartheta(T_2-T_1)}-1\big)e^{-\vartheta(T_1-t)}\sigma_t dB_t\\
&\qquad+\kappa_2^n\Delta_i^n\epsilon^2-\kappa_1^n\Delta_i^n\epsilon^1\Big),\\
D_i^n=&\big(\kappa_2^n\Delta_i^n\epsilon^2-\kappa_1^n\Delta_i^n\epsilon^1\big)\big(\kappa_2^n\Delta_i^n\epsilon^2-e^{-\vartheta_n(T_2-T_1)}\kappa_1^n\Delta_i^n\epsilon^1\big).
\end{aligned}$$

We proved that

$$\sum_{i=1}^{\lfloor t/\Delta_n\rfloor}\mathbb{E}\Big(\Delta_n^{-1/2}\big(e^{-\vartheta(T_2-T_1)}-e^{-\vartheta_n(T_2-T_1)}\big)\chi_i^n\frac{\widehat{\overline{\sigma}}^2_{n,t_{i-1}}-\overline{\sigma}^2_{t_{i-1}}}{\overline{\sigma}^2_{t_{i-1}}\widehat{\overline{\sigma}}^2_{n,t_{i-1}}}\Big|\mathcal{G}_{i-1}\Big)\overset{u.c.p.}{\to}0$$

in Section 2.4.5. Now, $\mathbb{E}\big(\big|A_i^n+B_i^n+C_i^n\big||\mathcal{G}_{i-1}\big)$ is of order $\Delta_n^{3/2}$, and $\mathbb{E}\big(\big|D_i^n\big||\mathcal{G}_{i-1}\big)$ is of order $\Delta_n^{2\beta}$, that is of order $\Delta_n^{3/2}$ as $\beta>3/4$. As $\sup_{i\in\mathcal{I}_n}\big|\widehat{\overline{\sigma}}^2_{n,t_{i-1}}-\overline{\sigma}^2_{t_{i-1}}\big|$ is, by Theorem 3.4, of order $\Delta_n^{\alpha/(2\alpha+1)}$, we have, for all $t\in[0,T]$,

$$\sum_{i=1}^{\lfloor t/\Delta_n\rfloor}\mathbb{E}\Big(\Delta_n^{-1/2}(\Delta_i^n Y^2-\Delta_i^n Y^1)(\Delta_i^n Y^2-e^{-\vartheta_n(T_2-T_1)}\Delta_i^n Y^1)\frac{\widehat{\overline{\sigma}}^2_{n,t_{i-1}}-\overline{\sigma}^2_{t_{i-1}}}{\overline{\sigma}^2_{t_{i-1}}\widehat{\overline{\sigma}}^2_{n,t_{i-1}}}\Big|\mathcal{G}_{i-1}\Big)\overset{u.c.p.}{\to}0. \quad (3.6)$$

Moreover,

$$
\begin{aligned}
&\mathbb{E}\Big(\Big(\Delta_n^{-1/2}(\Delta_i^n Y^2-\Delta_i^n Y^1)(\Delta_i^n Y^2-e^{-\vartheta_n(T_2-T_1)}\Delta_i^n Y^1)\frac{\widehat{\overline{\sigma}}^2_{n,t_{i-1}}-\overline{\sigma}^2_{t_{i-1}}}{\overline{\sigma}^2_{t_{i-1}}\widehat{\overline{\sigma}}^2_{n,t_{i-1}}}\Big)^2\Big|\mathcal{G}_{i-1}\Big)\\
=&\Delta_n^{-1}\mathbb{E}\Big(\big((\Delta_i^n Y^2-\Delta_i^n Y^1)(\Delta_i^n Y^2-e^{-\vartheta_n(T_2-T_1)}\Delta_i^n Y^1)\big)^2\big|\mathcal{G}_{i-1}\Big)\Big(\frac{\widehat{\overline{\sigma}}^2_{n,t_{i-1}}-\overline{\sigma}^2_{t_{i-1}}}{\overline{\sigma}^2_{t_{i-1}}\widehat{\overline{\sigma}}^2_{n,t_{i-1}}}\Big)^2\\
\leq&\Delta_n^{-1}\mathbb{E}\Big(\big((\Delta_i^n Y^2-\Delta_i^n Y^1)(\Delta_i^n Y^2-e^{-\vartheta_n(T_2-T_1)}\Delta_i^n Y^1)\big)^2\big|\mathcal{G}_{i-1}\Big)\frac{\big(\sup_{i\in\mathcal{I}_n}\big|\widehat{\overline{\sigma}}^2_{n,t_{i-1}}-\overline{\sigma}^2_{t_{i-1}}\big|\big)^2}{\tilde{c}^8}.
\end{aligned}
$$

As

$$
\mathbb{E}\Big(\big((\Delta_i^n Y^2-\Delta_i^n Y^1)(\Delta_i^n Y^2-e^{-\vartheta_n(T_2-T_1)}\Delta_i^n Y^1)\big)^2\big|\mathcal{G}_{i-1}\Big)
$$

is of order Δ_n^2 and $\sup_i\big|\widehat{\overline{\sigma}}^2_{n,t_{i-1}}-\overline{\sigma}^2_{t_{i-1}}\big|$ is of order $\Delta_n^{\alpha/(2\alpha+1)}$, there remains that

$$
\sum_{i\in\mathcal{I}_n}\mathbb{E}\Big(\Big(\Delta_n^{-1/2}(\Delta_i^n Y^2-\Delta_i^n Y^1)(\Delta_i^n Y^2-e^{-\vartheta_n(T_2-T_1)}\Delta_i^n Y^1)\frac{\widehat{\overline{\sigma}}^2_{n,t_{i-1}}-\overline{\sigma}^2_{t_{i-1}}}{\overline{\sigma}^2_{t_{i-1}}\widehat{\overline{\sigma}}^2_{n,t_{i-1}}}\Big)^2\Big|\mathcal{G}_{i-1}\Big)\lesssim\Delta_n^{\alpha/(2\alpha+1)},
$$

which converges to 0 in probability. With this result and (3.6), by Lemma 3.4 in [49], we conclude that S''_n converges to 0 in probability, which gives the expected result.

Second step

As in Section 2.4.5, using the previous step we have that

$$
\Delta_n^{1/2}\Big(\sum_{i\in\mathcal{I}_n}\big(\widetilde{\ell}\big(\vartheta_n,\widehat{\overline{\sigma}}^2_n\big)^i-\widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}}\big)+\Delta_n^{-1}\widetilde{I}_{\vartheta,\sigma,\overline{\sigma}}(\vartheta_n-\vartheta)\Big)\tag{3.7}
$$

converges to 0 in probability, which remains true if we replace the deterministic sequence ϑ_n by the discretized version of $\hat{\vartheta}_{2,n}$.

The next step is to prove that

$$
\Delta_n\sum_{i\in\mathcal{I}_n}\Big(\widetilde{\ell}\big(\hat{\vartheta}_{2,n},\widehat{\overline{\sigma}}^2_n\big)^i\Big)^2\to\widetilde{I}_{\vartheta,\sigma,\overline{\sigma}}\tag{3.8}
$$

in probability. We once again refer to Section 2.4.5 where this was done with no model errors;

it remains true, as the term with no error is predominant. We then have

$$
\begin{aligned}
\Delta_n^{-1/2}\big(\tilde{\vartheta}_{2,n}-\vartheta\big)\widetilde{I}_{\vartheta,\sigma,\overline{\sigma}} =&\Delta_n^{-1/2}(\hat{\vartheta}_{2,n}-\vartheta)\widetilde{I}_{\vartheta,\sigma,\overline{\sigma}}+\Delta_n^{-1/2}\frac{\Delta_n\widetilde{I}_{\vartheta,\sigma,\overline{\sigma}}\sum_{i\in\mathcal{I}_n}\widetilde{\ell}\big(\hat{\vartheta}_{2,n},\widehat{\overline{\sigma}}^2_n\big)^i}{\Delta_n\sum_{i\in\mathcal{I}_n}\big(\widetilde{\ell}\big(\hat{\vartheta}_{2,n},\widehat{\overline{\sigma}}^2_n\big)^i\big)^2}\\
=&\Delta_n^{-1/2}(\hat{\vartheta}_{2,n}-\vartheta)\widetilde{I}_{\vartheta,\sigma,\overline{\sigma}}+\Delta_n^{-1/2}\Delta_n\sum_{i\in\mathcal{I}_n}\widetilde{\ell}\big(\hat{\vartheta}_{2,n},\widehat{\overline{\sigma}}^2_n\big)^i\\
&+\Delta_n^{-1/2}\Delta_n\widetilde{I}_{\vartheta,\sigma,\overline{\sigma}}\sum_{i\in\mathcal{I}_n}\widetilde{\ell}\big(\hat{\vartheta}_{2,n},\widehat{\overline{\sigma}}^2_n\big)^i\Big(\frac{1}{\Delta_n\sum_{i\in\mathcal{I}_n}\big(\widetilde{\ell}\big(\hat{\vartheta}_{2,n},\widehat{\overline{\sigma}}^2_n\big)^i\big)^2}-\frac{1}{\widetilde{I}_{\vartheta,\sigma,\overline{\sigma}}}\Big)\\
=&\Delta_n^{1/2}\sum_{i\in\mathcal{I}_n}\widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}}+o_{\mathbb{P}}(1).
\end{aligned}
$$

because of the convergence in probability of (3.7) towards 0 and the convergence (3.8).
In Section 2.4.5 we proved that $\Delta_n^{1/2}\big(\widetilde{I}_{\vartheta,\sigma,\overline{\sigma}}\big)^{-1/2}\sum_{i\in\mathcal{I}_n}\widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}}$ converges stably in law to a random variable defined on $(\widetilde{\Omega},\widetilde{\mathcal{F}},\widetilde{\mathbb{P}})$, which conditionally to $\mathcal{F}$ is Gaussian with variance $\big(\widetilde{I}_{\vartheta,\sigma,\overline{\sigma}}\big)^{-1}$. Therefore, to end the proof, there only remains to show that

$$
\sum_{i\in\mathcal{I}_n}\Delta_n^{1/2}\big(\widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}}-\widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}}\big)
$$

converges to 0 in probability. To do so, write $\Delta_n^{1/2}\widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}}-\Delta_n^{1/2}\widetilde{\ell}^i_{\vartheta,\sigma,\overline{\sigma}}$ as

$$
\frac{e^{-\vartheta(T_2-T_1)}(T_2-T_1)}{\big(1-e^{-\vartheta(T_2-T_1)}\big)^3}\Delta_n^{1/2}(A_i^n+B_i^n),
$$

where A_i^n is

$$
\begin{aligned}
&\big((\Delta_i^nY^2-\Delta_i^nY^1)(\Delta_i^nY^2-e^{-\vartheta(T_2-T_1)}\Delta_i^nY^1)-(\Delta_i^nX^2-\Delta_i^nX^1)(\Delta_i^nX^2-e^{-\vartheta(T_2-T_1)}\Delta_i^nX^1)\big)\\
&\qquad\times\frac{1}{\Delta_n\overline{\sigma}^2_{t_{i-1}}}
\end{aligned}
$$

and B_i^n is

$$
\begin{aligned}
&\big((\Delta_i^nY^2-\Delta_i^nY^1)(\Delta_i^nY^2-e^{-\vartheta(T_2-T_1)}\Delta_i^nY^1)-(\Delta_i^nX^2-\Delta_i^nX^1)(\Delta_i^nX^2-e^{-\vartheta(T_2-T_1)}\Delta_i^nX^1)\big)\\
&\qquad\times\Big(\frac{1}{\Delta_n\int_{t_{i-1}}^{t_i}\overline{\sigma}_tdt}-\frac{1}{\Delta_n\overline{\sigma}^2_{t_{i-1}}}\Big).
\end{aligned}
$$

For $t\in[0,T]$, we prove that $\sum_{i=1}^{\lfloor t/\Delta_n\rfloor}\Delta_n^{1/2}A_i^n\to 0$ in probability using Lemma 3.4 in [49], by checking conditions (3.41) and (3.42) of that lemma. We have $\mathbb{E}(\Delta_n^{1/2}A_i^n|\mathcal{G}_{i-1})=\mathbb{E}(\mathbb{E}(\Delta_n^{1/2}A_i^n|\mathcal{H}_{i-1})|\mathcal{G}_{i-1})=\Delta_n^{-1/2}\kappa_1^n\kappa_2^ne^{-\vartheta(T_2-T_1)}\epsilon^1_{i-1}\epsilon^2_{i-1}\frac{1}{\overline{\sigma}^2_{t_{i-1}}}$ plus some terms of lower order, linked to the drift process. It follows that

$$
\sum_{i=1}^{\lfloor t/\Delta_n\rfloor}\mathbb{E}\big(\big|\mathbb{E}(\Delta_n^{1/2}A_i^n|\mathcal{G}_{i-1})\big||\mathcal{G}_{i-1}\big)\lesssim\Delta_n^{2\beta-3/2},
$$

which converges to 0 in probability because we assumed that $\beta > 3/4$; this gives condition (3.41). Then

$$\mathbb{E}\Big[\big((\Delta_i^n Y^2 - \Delta_i^n Y^1)(\Delta_i^n Y^2 - e^{-\vartheta(T_2-T_1)}\Delta_i^n Y^1) \\ - (\Delta_i^n X^2 - \Delta_i^n X^1)(\Delta_i^n X^2 - e^{-\vartheta(T_2-T_1)}\Delta_i^n X^1)\big)^2 \Big| \mathcal{G}_{i-1}\Big]$$

has order $\Delta_n^{4\beta\wedge(2\beta+1)}$ and $\frac{\Delta_n}{\Delta_n^2 \bar{\sigma}_{t_{i-1}}^4} \leq \frac{\Delta_n^{-1}}{\tilde{c}^4}$, so that $\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \mathbb{E}\big(\Delta_n (A_i^n)^2 | \mathcal{G}_{i-1}\big) \lesssim \Delta_n^{2\beta-1}$, which goes to 0 as $n \to \infty$; condition (3.42) is proved. We conclude that $\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \Delta_n^{1/2} A_i^n \to 0$ in probability. Among others, we have, for $t = T$, that $\sum_{i=1}^{n} \Delta_n^{1/2} A_i^n \to 0$ in probability. As we did in the very end of Section 2.4.5, we can get that $\sum_{i \in \mathcal{I}_n} \Delta_n^{1/2} A_i^n \to 0$ in probability as well.

To prove that $\sum_{i \in \mathcal{I}_n} \Delta_n^{1/2} B_i^n \to 0$ in probability, we just have to reproduce the calculus of the term S_n'' in Step 1 of Section 2.4.5 or of the current section. This ends the proof.

3.5 Appendices

3.5.1 Technical lemmas

Lemma 3.6.1. *Let $0 < \alpha < \beta < \gamma < \delta$. The functions $f : x \mapsto -e^{(\beta-\alpha)x}\frac{e^{-\gamma x}-e^{-\beta x}}{e^{-\gamma x}-e^{-\alpha x}}$ and $g : x \mapsto \frac{e^{-\delta x}-e^{-\gamma x}}{e^{-\beta x}-e^{-\alpha x}}$, defined on $(0, +\infty)$, are decreasing.*

We have

$$f(x) = -e^{(\beta-\alpha)x}\frac{e^{-(\beta-\gamma)x} - 1}{e^{-(\alpha-\gamma)x} - 1} = -e^{(\beta-\alpha)x}\frac{1 - e^{-(\gamma-\beta)x}}{e^{-(\alpha-\beta)x} - e^{-(\gamma-\beta)x}} = -\frac{e^{-(\gamma-\beta)x} - 1}{e^{-(\gamma-\alpha)x} - 1},$$

which, by Lemma 2.4.2, is the opposite of the inverse of a decreasing function; it is thus decreasing.

Then

$$g(x) = \frac{e^{-\delta x} - e^{-\gamma x}}{e^{-\gamma x} - e^{-\beta x}}\frac{e^{-\gamma x} - e^{-\beta x}}{e^{-\beta x} - e^{-\alpha x}},$$

which is, by Lemma 2.4.2, the product of two positive increasing functions. It is therefore positive and decreasing.

Lemma 3.6.2. *Work under Assumptions 2.1 and 2.2. Then, for all $t \in [0,T]$, the convergences in probability*

$$\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \Delta_n^{-1}\mathbb{E}\Big(\int_{t_{i-1}}^{t_i}\Big(\int_{t_{i-1}}^{t}\dot\sigma_s d\dot W_s\Big)^2\tilde\sigma_t^2 dt\Big|\mathcal{G}_{i-1}\Big) \to \frac{1}{2}\int_0^t \dot\sigma_s^2\tilde\sigma_s^2 ds$$

and

$$\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \Delta_n^{-1}\mathbb{E}\Big(\int_{t_{i-1}}^{t_i}\Big(\int_{t_{i-1}}^{t}\dot\sigma_s d\dot W_s\Big)\Big(\int_{t_{i-1}}^{t}\tilde\sigma_s d\tilde W_s\Big)\tilde\sigma_t^2 dt\Big|\mathcal{G}_{i-1}\Big) \to 0$$

where $\dot\sigma_t d\dot W_t$ and $\tilde\sigma_t d\tilde W_t$ can both stand for either $e^{-\vartheta(T_j-t)}\sigma_t dB_t$ or $\overline{\sigma}_t d\overline{B}_t$, for any $j = 1,\dots,d$, hold in probability, as $n \to \infty$.

As a preliminary remark for the proof, notice that $\mathbb{E}(|\dot\sigma_t - \dot\sigma_s|^2 + |\tilde\sigma_t - \tilde\sigma_s|^2) \lesssim \Delta_n$ as soon as $|t-s| < 1$.

First we prove that the first convergence holds: by the integration by parts formula,

$$\mathbb{E}\Big(\int_{t_{i-1}}^{t_i}\Big(\int_{t_{i-1}}^{t}\dot\sigma_s d\dot W_s\Big)^2\tilde\sigma_t^2 dt\Big|\mathcal{G}_{i-1}\Big) = A_i^n - B_i^n,$$

where

$$A_i^n = \mathbb{E}\Big(\int_{t_{i-1}}^{t_i}\tilde\sigma_t^2 dt\Big(\int_{t_{i-1}}^{t_i}\dot\sigma_t d\dot W_t\Big)^2\Big|\mathcal{G}_{i-1}\Big)$$

and

$$B_i^n = \mathbb{E}\Big(\int_{t_{i-1}}^{t_i}\Big(\int_{t_{i-1}}^{t}\tilde\sigma_s^2 ds\Big)\dot\sigma_t^2 dt\Big|\mathcal{G}_{i-1}\Big).$$

We have

$$\begin{aligned}\mathbb{E}\Big(B_i^n - \frac{\Delta_n^2}{2}\dot\sigma_{t_{i-1}}^2\tilde\sigma_{t_{i-1}}^2\Big) =& \mathbb{E}\Big(\mathbb{E}\Big(\int_{t_{i-1}}^{t_i}\Big(\int_{t_{i-1}}^{t}(\tilde\sigma_s^2 - \tilde\sigma_{t_{i-1}}^2)ds\Big)\dot\sigma_t^2 dt\Big|\mathcal{G}_{i-1}\Big)\Big)\\ &+ \mathbb{E}\Big(\tilde\sigma_{t_{i-1}}^2\mathbb{E}\Big(\int_{t_{i-1}}^{t_i}(t - t_{i-1})(\dot\sigma_t^2 - \dot\sigma_{t_{i-1}})dt\Big|\mathcal{G}_{i-1}\Big)\Big).\end{aligned}$$

By localization, we may assume that $\dot\sigma$ and $\tilde\sigma$ are bounded by some positive constant M_σ. Then, as soon as $\Delta_n < 1$,

$$\begin{aligned}\mathbb{E}\Big(\Big|\mathbb{E}\Big(\int_{t_{i-1}}^{t_i}\Big(\int_{t_{i-1}}^{t}(\tilde\sigma_s^2 - \tilde\sigma_{t_{i-1}}^2)ds\Big)\dot\sigma_t^2 dt\Big|\mathcal{G}_{i-1}\Big)\Big|\Big) &\le M_\Sigma^2\int_{t_{i-1}}^{t_i}\int_{t_{i-1}}^{t}\mathbb{E}\big(|\tilde\sigma_s^2 - \tilde\sigma_{t_{i-1}}^2|\big)dsdt\\ &\le M_\Sigma^2\int_{t_{i-1}}^{t_i}\int_{t_{i-1}}^{t}\sqrt{\mathbb{E}\big(|\tilde\sigma_s^2 - \tilde\sigma_{t_{i-1}}^2|^2\big)}dsdt\\ &\lesssim \Delta_n^{5/2},\end{aligned}$$

using Jensen inequality for $x \mapsto \sqrt{x}$. In the same way, we have

$$\mathbb{E}\Big(\tilde{\sigma}^2_{t_{i-1}}\mathbb{E}\Big(\int_{t_{i-1}}^{t_i}(t-t_{i-1})(\dot{\sigma}^2_t-\dot{\sigma}^2_{t_{i-1}})dt\Big|\mathcal{G}_{i-1}\Big)\Big) \lesssim \Delta_n^{5/2},$$

so that

$$\sum_{i=1}^{\lfloor t/\Delta_n\rfloor}\Delta_n^{-1}B_i^n = \sum_{i=1}^{\lfloor t/\Delta_n\rfloor}\Delta_n^{-1}\frac{\Delta_n^2}{2}\dot{\sigma}^2_{t_{i-1}}\tilde{\sigma}^2_{t_{i-1}} + o_{\mathbb{P}}(1) = \frac{1}{2}\int_0^t \dot{\sigma}^2_s\tilde{\sigma}^2_s ds + o_{\mathbb{P}}(1).$$

Now,

$$\begin{aligned}
A_i^n =&\tilde{\sigma}^2_{t_{i-1}}\Delta_n\mathbb{E}\Big(\int_{t_{i-1}}^{t_i}\dot{\sigma}^2_t dt\Big|\mathcal{G}_{i-1}\Big) + \mathbb{E}\Big(\int_{t_{i-1}}^{t_i}(\tilde{\sigma}^2_t-\tilde{\sigma}^2_{t_{i-1}})dt\Big(\int_{t_{i-1}}^{t_i}\dot{\sigma}_t d\dot{W}_t\Big)^2\Big|\mathcal{G}_{i-1}\Big)\\
=&\tilde{\sigma}^2_{t_{i-1}}\dot{\sigma}^2_{t_{i-1}}\Delta_n^2 + \tilde{\sigma}^2_{t_{i-1}}\Delta_n\mathbb{E}\Big(\int_{t_{i-1}}^{t_i}(\dot{\sigma}^2_t-\dot{\sigma}^2_{t_{i-1}})dt\Big|\mathcal{G}_{i-1}\Big)\\
&+\mathbb{E}\Big(\int_{t_{i-1}}^{t_i}(\tilde{\sigma}^2_t-\tilde{\sigma}^2_{t_{i-1}})dt\Big(\int_{t_{i-1}}^{t_i}\dot{\sigma}_t d\dot{W}_t\Big)^2\Big|\mathcal{G}_{i-1}\Big).
\end{aligned}$$

We have

$$\sum_{i=1}^{\lfloor t/\Delta_n\rfloor}\Delta_n^{-1}\tilde{\sigma}^2_{t_{i-1}}\dot{\sigma}^2_{t_{i-1}}\Delta_n^2 \to \int_0^t \dot{\sigma}^2_s\tilde{\sigma}^2_s ds,$$

in probability, and

$$\mathbb{E}\Big(\Big|\tilde{\sigma}^2_{t_{i-1}}\Delta_n\mathbb{E}\Big(\int_{t_{i-1}}^{t_i}(\dot{\sigma}^2_t-\dot{\sigma}^2_{t_{i-1}})dt\Big|\mathcal{G}_{i-1}\Big)\Big|\Big) \le M_\Sigma^2\Delta_n\int_{t_{i-1}}^{t_i}\sqrt{\mathbb{E}\big(\dot{\sigma}^2_t-\dot{\sigma}^2_{t_{i-1}})^2\big)}dt \lesssim \Delta_n^{5/2}.$$

We care for the last term with Cauchy-Schwarz inequality as follows:

$$\begin{aligned}
&\mathbb{E}\Big(\Big|\mathbb{E}\Big(\int_{t_{i-1}}^{t_i}(\tilde{\sigma}^2_t-\tilde{\sigma}^2_{t_{i-1}})dt\Big(\int_{t_{i-1}}^{t_i}\dot{\sigma}_t d\dot{W}_t\Big)^2\Big|\mathcal{G}_{i-1}\Big)\Big|\Big)\\
\le&\sqrt{\mathbb{E}\Big(\Big(\int_{t_{i-1}}^{t_i}(\tilde{\sigma}^2_t-\tilde{\sigma}^2_{t_{i-1}})dt\Big)^2\Big)\mathbb{E}\Big(\Big(\int_{t_{i-1}}^{t_i}\dot{\sigma}_t d\dot{W}_t\Big)^4\Big)}.
\end{aligned}$$

The second expectation in the product is of order Δ_n^2 by Burkholder-Davis-Gundy inequality, while, by Jensen inequality,

$$\mathbb{E}\Big(\Big(\int_{t_{i-1}}^{t_i}(\tilde{\sigma}^2_t-\tilde{\sigma}^2_{t_{i-1}})dt\Big)^2\Big) \le \Delta_n\int_{t_{i-1}}^{t_i}\mathbb{E}\big((\tilde{\sigma}^2_t-\tilde{\sigma}^2_{t_{i-1}})^2\big)dt \lesssim \Delta_n^3.$$

There remains

$$\mathbb{E}\Big(\Big|\mathbb{E}\Big(\int_{t_{i-1}}^{t_i}(\tilde{\sigma}^2_t-\tilde{\sigma}^2_{t_{i-1}})dt\Big(\int_{t_{i-1}}^{t_i}\dot{\sigma}_t d\dot{W}_t\Big)^2\Big|\mathcal{G}_{i-1}\Big)\Big|\Big) \lesssim \Delta_n^{5/2},$$

so that

$$\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \Delta_n^{-1} A_i^n = \int_0^t \dot{\sigma}_s^2 \tilde{\sigma}_s^2 ds + o_{\mathbb{P}}(1),$$

and finally,

$$\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \Delta_n^{-1} \mathbb{E}\Big(\int_{t_{i-1}}^{t_i} \Big(\int_{t_{i-1}}^{t} \dot{\sigma}_s d\dot{W}_s \Big)^2 \tilde{\sigma}_t^2 dt \Big| \mathcal{G}_{i-1} \Big) = \sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \Delta_n^{-1} (A_i^n - B_i^n) \to \frac{1}{2} \int_0^t \dot{\sigma}_s^2 \tilde{\sigma}_s^2 ds$$

in probability. This proves the first part of the lemma.

Now we establish the second convergence: by the integration by parts formula, the product $\int_{t_{i-1}}^{t_i} \Big(\int_{t_{i-1}}^{t} \dot{\sigma}_s d\dot{W}_s \Big) \Big(\int_{t_{i-1}}^{t} \tilde{\sigma}_s d\tilde{W}_s \Big) \tilde{\sigma}_t^2 dt$ is equal to the sum of some local martingale and $\int_{t_{i-1}}^{t_i} \tilde{\sigma}_t^2 dt \int_{t_{i-1}}^{t_i} \dot{\sigma}_t d\dot{W}_t \int_{t_{i-1}}^{t_i} \tilde{\sigma}_t d\tilde{W}_t$, so that

$$\begin{aligned}
&\mathbb{E}\Big(\int_{t_{i-1}}^{t_i} \Big(\int_{t_{i-1}}^{t} \dot{\sigma}_s d\dot{W}_s \Big) \Big(\int_{t_{i-1}}^{t} \tilde{\sigma}_s d\tilde{W}_s \Big) \tilde{\sigma}_t^2 dt \Big| \mathcal{G}_{i-1} \Big) \\
=& \Delta_n \tilde{\sigma}_{t_{i-1}}^2 \mathbb{E}\Big(\int_{t_{i-1}}^{t_i} \dot{\sigma}_t d\dot{W}_t \int_{t_{i-1}}^{t_i} \tilde{\sigma}_t d\tilde{W}_t \Big| \mathcal{G}_{i-1} \Big) \\
&+ \mathbb{E}\Big(\int_{t_{i-1}}^{t_i} (\tilde{\sigma}_t^2 - \tilde{\sigma}_{t_{i-1}}^2) dt \int_{t_{i-1}}^{t_i} \dot{\sigma}_t d\dot{W}_t \int_{t_{i-1}}^{t_i} \tilde{\sigma}_t d\tilde{W}_t \Big| \mathcal{G}_{i-1} \Big).
\end{aligned}$$

The first expectation in the RHS is zero, and by Cauchy-Schwarz inequality,

$$\begin{aligned}
&\mathbb{E}\Big(\Big| \mathbb{E}\Big(\int_{t_{i-1}}^{t_i} (\tilde{\sigma}_t^2 - \tilde{\sigma}_{t_{i-1}}^2) dt \int_{t_{i-1}}^{t_i} \dot{\sigma}_t d\dot{W}_t \int_{t_{i-1}}^{t_i} \tilde{\sigma}_t d\tilde{W}_t \Big| \mathcal{G}_{i-1} \Big) \Big| \Big) \\
\leq& \sqrt{ \mathbb{E}\Big(\Big(\int_{t_{i-1}}^{t_i} (\tilde{\sigma}_t^2 - \tilde{\sigma}_{t_{i-1}}^2) dt \Big)^2 \Big) \mathbb{E}\Big(\Big(\int_{t_{i-1}}^{t_i} \dot{\sigma}_t d\dot{W}_t \int_{t_{i-1}}^{t_i} \tilde{\sigma}_t d\tilde{W}_t \Big)^2 \Big) }.
\end{aligned}$$

We have already noticed that the expectation $\mathbb{E}\Big(\Big(\int_{t_{i-1}}^{t_i} (\tilde{\sigma}_t^2 - \tilde{\sigma}_{t_{i-1}}^2) dt \Big)^2 \Big)$ has order Δ_n^3, and that $\mathbb{E}\Big(\Big(\int_{t_{i-1}}^{t_i} \dot{\sigma}_t d\dot{W}_t \int_{t_{i-1}}^{t_i} \tilde{\sigma}_t d\tilde{W}_t \Big)^2 \Big)$ has Δ_n^2. It follows that

$$\mathbb{E}\Big(\Big| \mathbb{E}\Big(\int_{t_{i-1}}^{t_i} (\tilde{\sigma}_t^2 - \tilde{\sigma}_{t_{i-1}}^2) dt \int_{t_{i-1}}^{t_i} \dot{\sigma}_t d\dot{W}_t \int_{t_{i-1}}^{t_i} \tilde{\sigma}_t d\tilde{W}_t \Big| \mathcal{G}_{i-1} \Big) \Big| \Big) \lesssim \Delta_n^{5/2},$$

so that

$$\sum_{i=1}^{\lfloor t/\Delta_n \rfloor} \Delta_n^{-1} \mathbb{E}\Big(\int_{t_{i-1}}^{t_i} \Big(\int_{t_{i-1}}^{t} \dot{\sigma}_s d\dot{W}_s \Big) \Big(\int_{t_{i-1}}^{t} \tilde{\sigma}_s d\tilde{W}_s \Big) \tilde{\sigma}_t^2 dt \Big| \mathcal{G}_{i-1} \Big) \to 0$$

in probability, which ends the proof.

3.5.2 Plots of nonparametric estimators with model errors

Plots with $\beta = 0.55$ and $n = 1000$

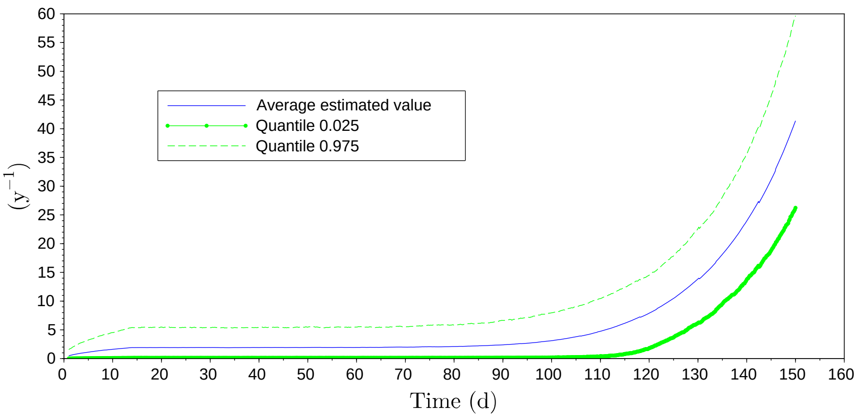

Figure 3.9 – Quantiles for the square of the equivalent volatility with 2 processes, $\beta = 0.55$ and $n = 1000$

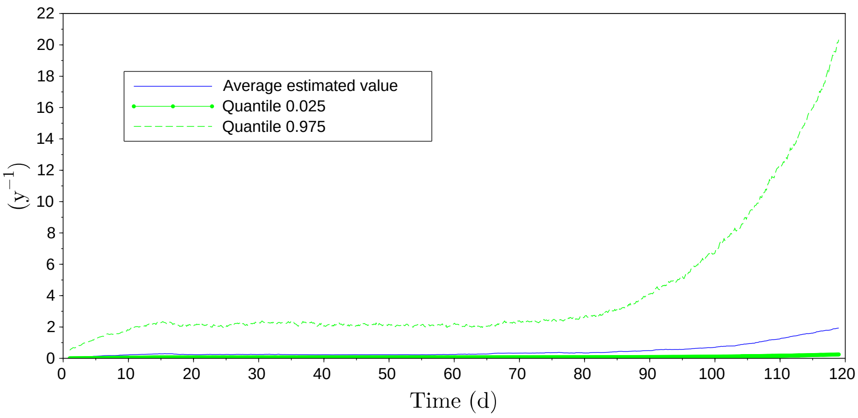

Figure 3.10 – Quantiles for the square of the equivalent volatility with 3 processes and $\beta = 0.55$ and $n = 1000$

Plots with $\beta = 0.625$ and $n = 100$

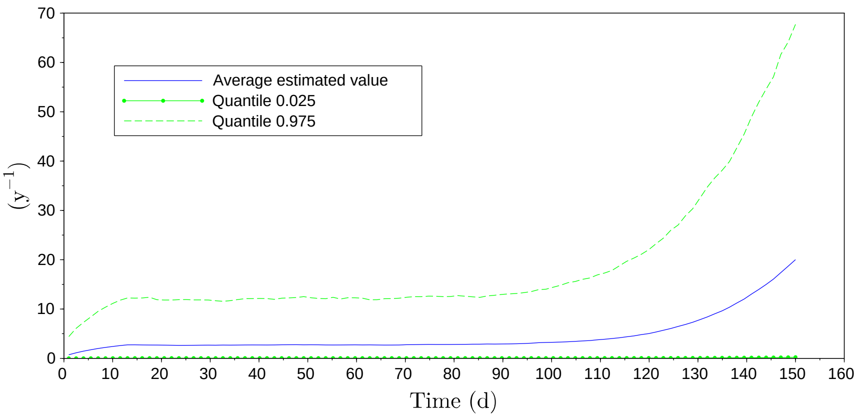

Figure 3.11 – Quantiles for the square of the equivalent volatility with 2 processes and $\beta = 0.625$

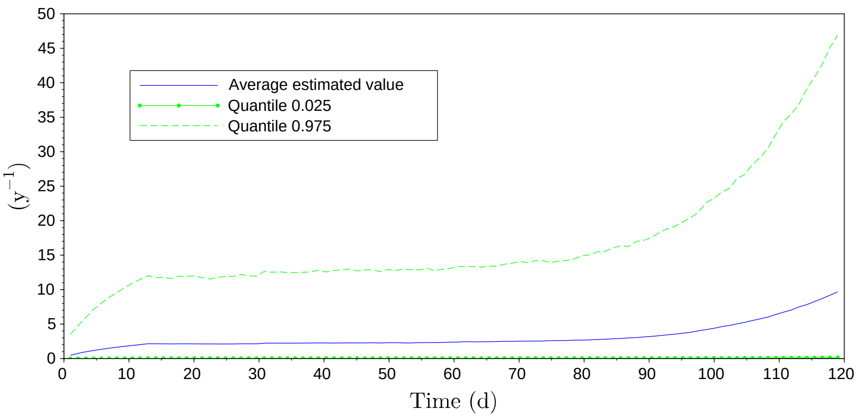

Figure 3.12 – Quantiles for the square of the equivalent volatility with 3 processes and $\beta = 0.6255$

Chapitre 4

An optimal trading problem in intraday electricity markets

Abstract

We consider the problem of optimal trading for a power producer in the context of intraday electricity markets. The aim is to minimize the imbalance cost induced by the random residual demand in electricity, i.e. the consumption from the clients minus the production from renewable energy. For a simple linear price impact model and a quadratic criterion, we explicitly obtain approximate optimal strategies in the intraday market and thermal power generation, and exhibit some remarkable properties of the trading rate. Furthermore, we study the case when there are jumps on the demand forecast and on the intraday price, typically due to error in the prediction of wind power generation. Finally, we solve the problem when taking into account delay constraints in thermal power production.

4.1 Introduction

The development of renewable energy sources in Europe as a response to global climate change has led to an increase of exchange in the intraday electricity markets. For instance, the exchanged volume on the European Energy Exchange (EEX) for Germany has grown from 2 TWh in 2008 to 25 TWh in 2013. This increase is mainly due to the level of forecasting error of wind production, which leads power producers owning a large share of wind production to turn more than ever to intraday markets in order to adjust their position and avoid penalties for their imbalances. The accuracy of forecasts for renewable power production from wind and solar may vary considerably depending on the agreggation level (local vs regional forecast) and the time horizon. For a complete survey on this problem, the reader can consult Giebel et al. [35], and may have in mind that the root mean square error (RMSE) of the error forecast for the production of a wind farm in six hours can reach 20% of its installed capacity. Many different intraday markets have been designed and are subject to different sets of regulation. But, in all cases, intraday markets offer power producer the possibility to buy or sell power for the next (say) 9 hours to 32 hours (case of the French electricity market of EpexSpot). These trades can occur after the closing of the day-ahead market or during the clearing phase of the day-ahead market. Moreover, there is a clear evidence that traders take the existence of some market impact into account. Indeed, for a given hour of delivery, the average volume sold or purchased in 2014 is of order of magnitude 340 MW while the average trade order volume is of approximate size 20 MW (source: Epexspot). This point indicates that traders split their sales or their purchases into small quantities to reduce their impact.

The problem of trading management in the intraday electricity market for a balancing purpose has already drawn the attention in the literature. Henriot [41] studied the problem of how the intraday market can help a power producer to deal with the wind production error forecast in a stylized discrete time model. In his model, the power producer is a wind producer who is trying to minimize her sourcing cost on the intraday market while maintaining a balance position between her forecast production and her sales. Henriot's model takes into account the impact of the wind power producer on the intraday price with a deterministic inverse demand function, and the intraday price is not a risk factor. The only risk factor comes from the error forecast of the wind production and its auto-correlation. Garnier and Madlener [30] studies the trade-off between entering into a deal in the intraday market right now and postponing it in a discrete time decision model where intraday prices follow a geometric Brownian model and wind production error forecast follows an arithmetic Brownian motion. In their framework, the power producer is supposed to have no impact on intraday prices. Liquidity risk is taken into account as a probability of not finding a counter-party at the next trading window.

In this chapter, we consider a power producer having at disposal some renewable energy sources (e.g. wind and solar), and thermal plants (e.g. coal, gas, oil, and nuclear sources), and who can buy/sell energy in the intraday markets. Her purpose is to minimize the imbalance cost, i.e. the cost induced by the difference between the demand of her clients minus the electricity produced and traded, plus the production and trading costs. In contrast with thermal power plants whose generation can be controlled, the power generated from renewable sources is subject to non controllable fluctuations or risks (wind speed, weather forecast) and is then considered here as a random factor just like the demand. We then call *the residual demand* the demand minus the energy generated by renewable energy. Thus, the problem of the power producer is to minimize the imbalance costs arising from her residual demand by relying both on her own controllable thermal assets and on the intraday market. As in [30], we assume that the power producer has access to a continuously updated forecast of the residual demand to be satisfied at terminal date T and that this forecast evolves randomly. Moreover, the intraday price for delivery at time T evolves also randomly and is correlated with the residual demand forecast. However, compared with [30], the intraday market can be used for optimization purposes. We develop a model that allows us to study how power producers can take advantage of the interaction between the dynamics of the residual demand forecast and the dynamics of the intraday prices.

Our model shares some links with optimal order execution problems, as introduced in the seminal paper by Almgren and Chriss [12], and then largely studied in the recent literature, see e.g. the survey paper [71]. In our context, the original feature with respect to this literature is the consideration of a random demand target and the possibility for the agent to use her thermal power production. This connection with optimal execution is fruitful in the sense that it allows us to take into account several features of intraday markets while maintaining the tractability of the model sufficiently high to allow analytical solutions. Hence, we take into account liquidity risk through a market impact, both permanent and temporary, on the electricity price generated by a power producer when trading in the intraday market. As in optimal execution problems, this impact is always in the adverse direction: when the producer sells, the price decreases and when she buys, the price increases. Our setting is a continuous-time decision problem representing the possibility for the producer to make a deal at each time she wants and not only at pre-specified windows. Moreover, it is general enough as it permits us to study the limiting cases of a pure retailer (no production function), a pure trader (no demand commitment) and an integrated player (player owning both clients and generation), small or large.

The main goal of this chapter is to derive analytical results, which provide explicit solutions for the (approximate) optimal control, hence giving enlightening economic interpretations of the optimal trading strategies. In order to achieve such analytical tractability, we have to make some simplifying assumptions on the dynamics of the price process and of the residual demand forecast, as well as on the cost function, assumed to be of quadratic form meaning a simple linear growth of the marginal cost of production with respect to the production level. We first consider a simple model for a continuous price process with linear impact, and demand forecast driven by an arithmetic Brownian motion, and neglect in a first step the delay of production when using thermal power plants. We then study an auxiliary control problem by relaxing the non-negativity constraint on the generation level, for which we are able to derive explicit solutions. The approximation error induced by this relaxation constraint is analyzed. In next steps, we consider more realistic situations and investigate two extensions: (i) On one hand, we incorporate the case where the residual demand forecast is subject to sudden changes, related to prediction error for wind or solar power production, which may be quite important due to the difficulties for estimating wind speed and forecasting weather, see [15]. This is formalized by jumps in the dynamics of the demand process, and consequently also on the price process. Again, we are able to obtain explicit solutions. Actually, the key tool in the derivation of all these analytical results is a suitable treatment of the linear-quadratic structure of our stochastic control problem. (ii) On the other hand, we introduce natural delay constraints in the production, and show how the optimal decision problem can be explicitly solved by a suitable reduction to a problem without delay.

Our (approximate) optimal trading strategies present some remarkable properties. When the intraday price process is a martingale, the optimal trading rate inherits the martingale property, which implies in particular that the net position of electricity shares has a constant growth rate on average. Moreover, the optimal strategy consists in making at each time the forecast marginal cost equal to the forecast intraday price. This property follows the common sense of intraday traders. Consequently, if the producer has made sales or purchases on the day-ahead such that her forecast marginal cost equals the day-ahead price and if the initial condition of the intraday price is the day-ahead price, thus, on average, the producer optimal trading rate is zero. This fact is no longer true when the demand forecast and the price follow processes with jumps. In this case, the optimal trading rate is a supermartingale or a submartingale depending on the relative probability and size of positive and negative jumps on the price process. For this reason, contrary to the case without jumps, the power producer may need to have a non-zero initial trading rate even if she has made sales or purchases on the day-ahead such that her forecast marginal cost equals the day-ahead price and if the initial condition of the intraday price is the day-ahead price. We also quantify explicitly the impact of delay in production on the trading strategies. When the price process is a martingale, the net inventory in electricity shares grows linearly on average, with a change of slope (which is smaller) at the time decision for the production.

The outline of the chapter is organized as follows. We formulate the optimal trading problem in Section 2. In Section 3, we study the optimal trading problem without delay. We first solve explicitly the auxiliary optimal execution problem, and then study the approximation on the solution to the original problem, by focusing in particular on the error asymptotics. We illustrate our results with some numerical tests and simulations. We extend in Section 4 our results to the case where jumps in demand forecast may arise. In Section 5, we show how the optimal trading problem with delay in production can be reduced to a problem without delay, and then leads to explicit solutions. Finally, the appendix collects the explicit derivations of our solutions, which are justified by verification theorems.

4.2 Problem formulation

We consider an agent on an intraday energy market, who is required to guarantee her equilibrium supply/demand for a given fixed time T : she has to satisfy the demand of her customers by purchase/sale of energy on the intraday market at time T and also by means of her thermal power generation. We denote by X_t the net position of sales/purchases of electricity at time $t \leq T$ for a delivery at terminal time T, assumed to be described by an absolutely continuous trajectory up to time T, and by $q_t = \dot{X}_t$ the trading rate: $q_t > 0$ means an instantaneous purchase of electricity, while $q_t < 0$ represents an instantaneous sale at time t :

$$X_t \quad = \quad X_0 + \int_0^t q_s ds, \quad 0 \leq t \leq T. \tag{4.1}$$

Given the trading rate, the transactions occur with a market price impact:

$$P_t(q) \quad = \quad \hat{P}_t + \int_0^t g(q_s) ds + f(q_t).$$

Here, $(\hat{P}_t)_t$ is the unaffected intraday electricity price process on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}, \mathbb{P})$, carrying some part of randomness of the market, and following the terminology in the seminal paper by Almgren and Chriss [12], the term $f(q_t)$ refers to the temporary price impact, while $\int_0^t g(q_s)ds$ describes the permanent price impact. The price $(\hat{P}_t)_t$ may be seen as a forward price, evolving in real time, for delivery at time T. Let us then denote by Y the intraday electricity price impacted by the past trading rate q of the agent, defined by:

$$Y_t \quad := \quad \hat{P}_t + \int_0^t g(q_s)ds.$$

We assume that Y_t is observable and quoted, which means actually that the agent is a large trader and electricity producer, whose actions directly impact the intraday electricity price. The case where the agent is a small producer can be also dealt with by simply considering a zero permanent impact function $g \equiv 0$. Notice that the transacted price is equal to the sum of the quoted price Y and the temporary price impact:

$$P_t(q) \quad = \quad Y_t + f(q_t). \tag{4.2}$$

The residual demand D_T is the consumption of clients of the agent minus the production from renewable energy at terminal date T, and we assume that the agent has access to a continuously updated forecast $(D_t)_t$ of the residual demand. The agent can use her thermal power production with a quantity ξ at cost $c(\xi)$ in order to match as close as possible the target demand D_T. In practice, generation of electricity cannot be obtained instantaneously and needs a delay to reach a required level of production. Hence, the decision to produce a quantity ξ should be taken at time $T - h$, where $h \in [0, T]$ is the delay. Thus, for a controlled trading rate $q = (q_t)_t \in \mathcal{A}$, the set of real-valued $\mathbb{F}$-adapted processes satisfying some integrability conditions to be precised later, a production quantity $\xi \in L^0_+(\mathcal{F}_{T-h})$, the set of nonnegative $\mathcal{F}_{T-h}$-measurable random variables, the total cost is:

$$\int_0^T q_t P_t(q)dt + C(D_T - X_T, \xi) \quad := \quad \int_0^T q_t P_t(q)dt + c(\xi) + \frac{\eta}{2}(D_T - X_T - \xi)^2. \tag{4.3}$$

The first term in (4.3) represents the total running cost arising from the trading in the intraday electricity market, and the last term, where $\eta > 0$, represents the quadratic penalization when the net position in sales/purchases of electricity $X_T + \xi$ (including the production quantity ξ at cost $c(\xi)$) at terminal date T does not fit the effective demand D_T. The objective of the agent is then to minimize over q and ξ the expected total cost:

$$\text{minimize over } q \in \mathcal{A},\ \xi \in L^0_+(\mathcal{F}_{T-h}) \qquad \mathbb{E}\Big[\int_0^T q_t P_t(q)dt + C(D_T - X_T, \xi)\Big]. \tag{4.4}$$

Remark 4.1. 1) The imbalance of the agent $(D_T - X_T - \xi)$ is penalized by the Transport System Operator (TSO) because if a producer generates less power than her demand, then the TSO has to buy the energy from another producer to insure that the total production of all producers is equal to the total demand of the electric system. When the producer generates too much power, she is not truly penalized, but this excess of energy is bought back by the

TSO at a price that is lower than the marginal cost of the producer. The penalization term in the objective function above is a simplification of the effective penalization process that can be found in real electricity markets. For instance, the penalization of imbalances in the French electricity market depends both on the sign of the imbalance of the electricity system and on the price of imbalances (see [1, chap 2., sec. 2.2.1]). Nevertheless, the positive coefficient η captures the main objective of the penalization process. The agent has no incentive of being either too long or too short.

2) On real markets, trading ends some time before the date of delivery, at which the agent has to ensure equilibrium (e.g. on the French electricity market, there is a delay of 45 minutes). We do not include that practical fact in our framework, by considering that the delay is null for the sake of clarity. There is no mathematical consequence: it is enough to have in mind that the delivery and production do not really take place at T, but at T plus some delay.

3) The larger is η, the stronger is the incentive for the agent to be as close as possible to the equilibrium supply/demand. At the limit, when η goes to infinity, the agent is formally constrained to fit supply and demand. However, the limiting problem when $\eta = \infty$ is not mathematically well-posed since such perfect equilibrium constraint is in general not achievable. Indeed, the demand at terminal date T is random, typically modelled via a Gaussian noise, and the inventory X which is of finite variation, may exceed or underperform with positive probability the demand D_T at terminal date T. Hence, in the scenario where $X_T > D_T$, and since by nature the production quantity ξ is nonnegative, it is not possible to realize the equilibrium $X_T + \xi = D_T$, even if there is no delay. In the sequel, we fix $\eta > 0$ (which may be large, but finite), and study the stochastic control problem (4.4).

4) The optimization problem (4.4) shares somes similarities with the optimal execution problem in limit order book studied in the seminal paper by Almgren and Chriss [12], and then extended by many authors in the literature, see e.g. the survey paper [71]. The main difference is that in the execution problem of equities, the target is to buy or sell a certain number of shares, i.e. lead X_T to a fixed constant (meaning formally that η goes to infinity) while in our intraday electricity markets context, the target is to realize the equilibrium with the random demand D_T, eventually with the help of production leverage ξ. However, in contrast with the case of constant target, it is not possible in presence of random target D_T to achieve perfectly the equilibrium, which justifies the introduction of the penalty factor η as pointed out above. □

The main aim of this chapter is to provide explicit (or at least approximate explicit) solutions to the optimization problem (4.4), which are easily interpreted from an economic point of view, and also allow to measure the impact of the various parameters of the model. In order to achieve this goal, we shall adapt our modeling as close as possible to the linear-quadratic framework of stochastic control, and make the following assumption: The energy production cost function is in the quadratic form:

$$c(x) = \frac{\beta}{2}x^2,$$

for some $\beta > 0$. Although simple, a quadratic cost function represents the increase of the marginal cost of production with the level of production.

Remark 4.2. *(Pure retailer)* In the limiting case when β goes to infinity, meaning an infinite cost of production, this corresponds to the framework where the agent never uses the production leverage and only trades in the intraday-market by solving the optimal execution problem:

$$\text{minimize over } q \in \mathcal{A} \qquad \mathbb{E}\Big[\int_0^T q_t P_t(q)dt + C(D_T - X_T, 0)\Big]. \tag{4.5}$$

□

As in Almgren and Chriss, we assume that the price impact (both permanent and temporary) is of linear form, i.e.

$$g(q) \;=\; \nu q, \qquad f(q) \;=\; \gamma q,$$

for some constants $\nu \geq 0$ and $\gamma > 0$. The unaffected intraday electricity price is taken as a Bachelier model:

$$\hat{P}_t \;=\; \hat{P}_0 + \sigma_0 W_t, \tag{4.6}$$

where W is a standard Brownian motion, and $\sigma_0 > 0$ is a positive constant. Such assumption might seem a shortcoming at first sight since it allows for negative values of the unaffected price. However, in practice, for our intraday execution problem within few hours, negative prices occur only with negligible probability. This issue has been addressed in several works, see for instance Footnote 8 in [12], the comments in [71], or [38]. The martingale assumption is also standard in the market impact literature since drift effects can often be ignored due to short trading horizon. The quoted price Y, impacted by the past trading rate $q \in \mathcal{A}$, is then governed by the dynamics:

$$dY_t \;=\; \nu q_t dt + \sigma_0 dW_t. \tag{4.7}$$

The dynamics of the residual demand forecast is given by

$$dD_t \;=\; \mu dt + \sigma_d dB_t, \tag{4.8}$$

where μ, σ_d are constants, with $\sigma_d > 0$, and B is a Brownian motion correlated with W : $d < W, B >_t = \rho dt$, $\rho \in [-1, 1]$.
From (4.2), one can then define the value function associated to the dynamic version of the optimal execution problem (4.4) by:

$$v(t, x, y, d) \;:=\; \inf_{q \in \mathcal{A}_t, \xi \in L^0_+(\mathcal{F}_{T-h})} J(t, x, y, d; q, \xi) \tag{4.9}$$

with

$$J(t, x, y, d; q, \xi) \;:=\; \mathbb{E}\Big[\int_t^T q_s(Y_s^{t,y} + \gamma q_s)ds + C(D_T^{t,d} - X_T^{t,x}, \xi)\Big], \tag{4.10}$$

for $(t,x,y,d) \in [0,T] \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}$, where $\mathcal{A}_t$ denotes the set of real-valued processes $q = (q_s)_{t \leq s \leq T}$ s.t. q_s is $\mathcal{F}_s$-adapted and $\mathbb{E}\big[\int_t^T q_s^2 ds\big] < \infty$, $D^{t,d}$ is the solution to (4.8) starting from d at time t, and given a control $q \in \mathcal{A}_t$, $Y^{t,y}$ denotes the solution to (4.7) starting from y at time t, and $X^{t,x}$ is the solution to (4.1) starting from x at t.
In a first step, we shall consider the case when there is no delay in the production, and then show in the last section of this chapter how to reduce the problem with delay to a no delay problem. We shall also study the case when there are jumps in the residual demand forecast.

4.3 Optimal execution without delay in production

In this section, we consider the case when there is no delay in production, i.e. $h = 0$. In this case, we notice that the optimization over q and ξ in (4.4) is done separately. Indeed, the production quantity $\xi \in L^0_+(\mathcal{F}_T)$ is chosen at the final date T, after the decision over the trading rate process $(q_t)_{t\in[0,T]}$ is achieved (leading to an inventory X_T). It is determined optimally through the optimization a.s. at T of the terminal cost $C(D_T - X_T, \xi)$, hence in feedback form by $\xi^*_T = \hat{\xi}^{tr+}(D_T - X_T)$ where

$$\begin{aligned} \hat{\xi}^{tr+}(d) &:= \arg\min_{\xi \geq 0} C(d,\xi) \;=\; \arg\min_{\xi \geq 0} \Big[\frac{\beta}{2}\xi^2 + \frac{\eta}{2}(d-\xi)^2\Big] \\ &= \frac{\eta}{\eta+\beta} d \mathbf{1}_{d \geq 0}, \end{aligned} \tag{4.11}$$

the notation "$tr+$" indicating that some truncation of the negative part has been performed. The value function of problem (4.9) may then be rewritten as

$$v(t,x,y,d) \;=\; \inf_{q \in \mathcal{A}_t} \mathbb{E}\Big[\int_t^T q_s(Y_s^{t,y} + \gamma q_s) ds + C^+(D_T^{t,d} - X_T^{t,x})\Big], \tag{4.12}$$

where

$$\begin{aligned} C^+(d) &:= C(d, \hat{\xi}^{tr+}(d)) \\ &= \frac{1}{2}\frac{\eta\beta}{\eta+\beta} d^2 \mathbf{1}_{d \geq 0} + \frac{\eta}{2} d^2 \mathbf{1}_{d<0}. \end{aligned} \tag{4.13}$$

and the optimal trading rate q^* is derived by solving (4.12).
Due to the indicator function in C^+, caused by the non-negativity constraint on the production quantity, there is no hope to get explicit solutions for the problem (4.12), i.e. solve explicitly the associated dynamic programming Hamilton-Jacobi-Bellman (HJB) equation. We shall then consider an auxiliary execution problem by relaxing the sign constraint on the production quantity, for which we are able to provide explicit solution. Next, we shall see how one can derive an approximate solution to the original problem in terms of this auxiliary explicit solution, and we evaluate the error and illustrate the quality of this approximation by numerical tests.

4.3.1 Auxiliary optimal execution problem

We consider the optimal execution problem with relaxation on the non-negativity constraint of the production leverage, and thus introduce the auxiliary value function

$$\tilde{v}(t,x,y,d) \quad := \quad \inf_{q\in\mathcal{A},\xi\in L^0(\mathcal{F}_T)} J(t,x,y,d;q,\xi),$$

for $(t,x,y,d) \in [0,T]\times\mathbb{R}\times\mathbb{R}\times\mathbb{R}$. By same arguments as for the derivation of (4.12), we have

$$\tilde{v}(t,x,y,d) \quad = \quad \inf_{q\in\mathcal{A}} \mathbb{E}\Big[\int_t^T q_s(Y_s^{t,y}+\gamma q_s)ds + \tilde{C}(D_T^{t,d}-X_T^{t,x})\Big], \tag{4.14}$$

where

$$\begin{aligned}
\hat{\xi}(d) \quad &:= \quad \arg\min_{\xi\in\mathbb{R}} C(d,\xi) \ = \ \frac{\eta}{\eta+\beta}d,\\
\tilde{C}(d) \quad &:= \quad C(d,\hat{\xi}(d)) \ = \ \frac{1}{2}\frac{\eta\beta}{\eta+\beta}d^2 \ =: \ \frac{1}{2}r(\eta,\beta)d^2.
\end{aligned} \tag{4.15}$$

The function in (4.15) can be interpreted as a reduced cost function. Because the production cost function and the penalization are both quadratic, they can be reduced to a single production function where the imbalances are internalized by the producer.
The auxiliary problem (4.14)-(4.15) can be interpreted as a situation where it would be necessary to increase the demand. Allowing for negative generation is equivalent to include the possibility either to increase the demand through a price signal or to sell at a negative price on the intraday market. Those two possibilities exist on electricity markets.
By exploiting the linear-quadratic structure of the stochastic control problem (4.14), we can obtain explicit solutions for this auxiliary problem.

Theorem 4.1. *The value function to (4.14) is explicitly equal to:*

$$\begin{aligned}
\tilde{v}(t,x,y,d) \quad = \quad & \frac{r(\eta,\beta)(\frac{\nu}{2}(T-t)+\gamma)}{(r(\eta,\beta)+\nu)(T-t)+2\gamma}\big((d-x)^2+2\mu(T-t)(d-x)\big)\\
&+ \frac{T-t}{(r(\eta,\beta)+\nu)(T-t)+2\gamma}\big(-\frac{y^2}{2}+r(\eta,\beta)\mu(T-t)y\big)\\
&+ \frac{r(\eta,\beta)(T-t)}{(r(\eta,\beta)+\nu)(T-t)+2\gamma}(d-x)y\\
&+ \gamma\frac{\sigma_0^2+\sigma_d^2r^2(\eta,\beta)-2\rho\sigma_0\sigma_d r(\eta,\beta)}{\big(r(\eta,\beta)+\nu\big)^2}\ln\Big(1+\frac{(r(\eta,\beta)+\nu)(T-t)}{2\gamma}\Big)\\
&+ \frac{\sigma_d^2 r(\eta,\beta)\nu+2\rho\sigma_0\sigma_d r(\eta,\beta)-\sigma_0^2}{2\big(r(\eta,\beta)+\nu\big)}(T-t)\\
&+ \frac{r(\eta,\beta)\mu^2(T-t)^2(\frac{\nu}{2}(T-t)+\gamma)}{(r(\eta,\beta)+\nu)(T-t)+2\gamma},
\end{aligned}$$

for $(t,x,y,d) \in [0,T] \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}$, *with an optimal trading rate given in feedback form by:*

$$\begin{aligned}\hat{q}_s &= \hat{q}\big(T-s, D_s^{t,d} - \hat{X}_s^{t,x,y,d}, \hat{Y}_s^{t,x,y,d}\big), \quad t \le s \le T \\ \hat{q}(t,d,y) &:= \frac{r(\eta,\beta)(\mu t + d) - y}{(r(\eta,\beta)+\nu)t + 2\gamma}.\end{aligned} \tag{4.16}$$

Here $(\hat{X}^{t,x,y,d}, \hat{Y}^{t,x,y,d}, D^{t,d})$ *denotes the solution to (4.1)-(4.7)-(4.8) when using the feedback control* $\hat{q}$, *and starting from* (x,y,d) *at time* t. *Finally, the optimal production leverage is given by:*

$$\hat{\xi}_T = \hat{\xi}(D_T^{t,d} - \hat{X}_T^{t,x,y,d}) = \frac{\eta}{\eta+\beta}\big(D_T^{t,d} - \hat{X}_T^{t,x,y,d}\big). \tag{4.17}$$

Skech of proof. We look for a candidate solution to (4.14) in the quadratic form:

$$\begin{aligned}\tilde{w}(t,x,y,d) &= A(T-t)(d-x)^2 + B(T-t)y^2 + F(T-t)(d-x)y \\ &\quad + G(T-t)(d-x) + H(T-t)y + K(T-t),\end{aligned}$$

for some deterministic functions A, B, F, G, H and K. If we plug this ansatz into the Hamilton-Jacobi-Bellman (HJB) equation associated to the stochastic control problem (4.14), we find that these deterministic functions should satisfy a system of Riccati equations, which can be explicitly solved. Then, by a classical verification argument, we check that this ansatz $\tilde{w}$ is indeed equal to $\tilde{v}$, with an optimal feedback control derived from the argmax in the HJB equation. The details of the proof are reported in Appendix. $\square$

Remark 4.3. (*Pure trader*) By sending β to infinity in the expression of the value function $\tilde{v}$ and of the optimal feedback control $\hat{q}$, and observing that $r(\eta,\beta)$ goes to η, we obtain the solution to the optimal execution problem (4.5) without leverage production:

$$\begin{aligned}v_{NP}(t,x,y,d) &:= \inf_{q\in\mathcal{A}} \mathbb{E}\Big[\int_t^T q_s(Y_s^{t,y} + \gamma q_s)ds + C(D_T^{t,d} - X_T^{t,x}, 0)\Big] \\ &= \frac{\eta(\frac{\nu}{2}(T-t)+\gamma)}{(\eta+\nu)(T-t)+2\gamma}\big((d-x)^2 + 2\mu(T-t)(d-x)\big) \\ &\quad + \frac{T-t}{(\eta+\nu)(T-t)+2\gamma}\Big(-\frac{y^2}{2} + \eta\mu(T-t)y\Big) \\ &\quad + \frac{\eta(T-t)}{(\eta+\nu)(T-t)+2\gamma}(d-x)y \\ &\quad + \gamma\frac{\sigma_0^2 + \sigma_d^2\eta^2 - 2\rho\sigma_0\sigma_d\eta}{\big(\eta+\nu\big)^2}\ln\Big(1 + \frac{(\eta+\nu)(T-t)}{2\gamma}\Big) \\ &\quad + \frac{\sigma_d^2\eta\nu + 2\rho\sigma_0\sigma_d\eta - \sigma_0^2}{2\big(\eta+\nu\big)}(T-t) \\ &\quad + \frac{\eta\mu^2(T-t)^2(\frac{\nu}{2}(T-t)+\gamma)}{(\eta+\nu)(T-t)+2\gamma},\end{aligned} \tag{4.18}$$

for $(t,x,y,d) \in [0,T] \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}$, with an optimal trading rate given in feedback form by:

$$\begin{aligned} \hat{q}_s^{NP} &= \hat{q}^{NP}\big(T-s, D_s^{t,d} - \hat{X}_s^{t,x,y,d}, \hat{Y}_s^{t,x,y,d}\big), \quad t \le s \le T \\ \hat{q}^{NP}(t,d,y) &:= \frac{\eta(\mu t + d) - y}{(\eta+\nu)t + 2\gamma}. \end{aligned}$$

□

Interpretation:

1. The optimal trading rate $\hat{q}_s$ at time $s \in [t,T]$, given in feedback form by (4.16), is decomposed in two terms: the first one

$$\frac{r(\eta,\beta)}{(r(\eta,\beta)+\nu)(T-t)+2\gamma}\big(\mu(T-s) + D_s^{t,d} - \hat{X}_s^{t,x,y,d}\big)$$

is related to the trading rate in order to follow the trend of the demand, and to the incentive to invest when the forecast of the residual demand is larger than the current inventory. The second term

$$-\frac{1}{(r(\eta,\beta)+\nu)(T-t)+2\gamma}\hat{Y}_s^{t,x,y,d}$$

represents the negative impact of the quoted price on the investment strategy: the higher the price is, the more the agent decreases her trading rate until she reaches negative value meaning a resale of electricity shares. These effects are weighted by the constant denominator term depending on the penalty factor η, the marginal cost production factor β, the temporary and permanent price impact parameters γ, ν, and the time to maturity $T-t$.

2. By introducing the marginal cost function: $c'(x) = \beta x$, and the process

$$\hat{\xi}_s \;:=\; \frac{\eta}{\eta+\beta}\big(D_s^{t,d} + \mu(T-s) - \hat{X}_s^{t,x,y,d} - \hat{q}_s(T-s)\big), \quad t \le s \le T,$$

which is interpreted as the forecast production for the final time T (recall expression (4.17) of the final production), we notice from the expression of the optimal trading rate that the following relation holds:

$$\hat{Y}_s^{t,x,y,d} + \nu\hat{q}_s(T-s) + 2\gamma\hat{q}_s \;=\; c'(\hat{\xi}_s), \quad t \le s \le T. \tag{4.19}$$

This relation means that at each time, the optimal trading rate is to make the forecast intraday price plus marginal temporary impact (left hand side), which can be seen as the marginal cost of electricity on the intraday market at time T, equal to the forecast marginal cost of production. Here, the instantaneous impact γ appears as a marginal cost of buying or selling, and the forecast at time s supposes that the optimal trading rate $\hat{q}_s$ is held constant between s and T. □

We complete the description of the optimal trading rate by pointing out a remarkable martingale property.

Proposition 4.1. *The optimal trading rate process* $(\hat{q}_s)_{t\leq s\leq T}$ *in (4.16) is a martingale.*

Proof. By applying Itô's formula to $\hat{q}_s = \hat{q}(T-s, D_s^{t,d} - \hat{X}_s^{t,x,y,d}, \hat{Y}_s^{t,x,y,d})$, $t \leq s \leq T$, and since $\hat{q}$ is linear in d and y, we have:

$$\begin{aligned} d\hat{q}_s &= \big[-\frac{\partial \hat{q}}{\partial t} + (\mu - \hat{q})\frac{\partial \hat{q}}{\partial d} + \nu\hat{q}\frac{\partial \hat{q}}{\partial y}\big](T-s, D_s^{t,d} - \hat{X}_s^{t,x,y,d}, \hat{Y}_s^{t,x,y,d})ds \\ &\quad + \frac{\partial \hat{q}}{\partial d}(T-s, D_s^{t,d} - \hat{X}_s^{t,x,y,d}, \hat{Y}_s^{t,x,y,d})\sigma_d dB_s \\ &\quad + \frac{\partial \hat{q}}{\partial y}(T-s, D_s^{t,d} - \hat{X}_s^{t,x,y,d}, \hat{Y}_s^{t,x,y,d})\sigma_0 dW_s, \end{aligned}$$

from the dynamics (4.1), (4.8), and (4.7) of $\hat{X}^{t,x,y,d}$, $D^{t,d}$ and $\hat{Y}^{t,x,y,d}$. Now, from the explicit expression of the function $\hat{q}(t,y,d)$, we see that

$$-\frac{\partial \hat{q}}{\partial t} + (\mu - \hat{q})\frac{\partial \hat{q}}{\partial d} + \nu\hat{q}\frac{\partial \hat{q}}{\partial y} = 0,$$

and so:

$$d\hat{q}_s = \frac{r(\eta,\beta)\sigma_d}{(r(\eta,\beta)+\nu)(T-s)+2\gamma}dB_s - \frac{\sigma_0}{(r(\eta,\beta)+\nu)(T-s)+2\gamma}dW_s, \tag{4.20}$$

which shows the required martingale property. □

Remark 4.4. Recall that in the classical optimal execution problem as studied in [12], the optimal trading rate is constant. We retrieve this result in their framework which corresponds to the case where $\sigma_d = 0$ (constant demand target), $\beta = \infty$ (there is no production), and $\eta = \infty$ (constraint to lead X_T to the fixed target), see Remark 4.1 **4)**. Indeed, in these limiting regimes, we see from (4.20) that $d\hat{q}_s = 0$, meaning that $\{\hat{q}_s, t \leq s \leq T\}$ is constant. In our framework, this is generalized to the martingale property of the optimal trading rate process, which implies that the optimal inventory $\{\hat{X}_s^{t,x,y,d}, t \leq s \leq T\}$ has a constant growth rate in mean, i.e. $\frac{d\mathbb{E}[\hat{X}_s^{t,x,y,d}]}{ds}$ is constant equal to the initial trading rate at time t given by $\hat{q}(T-t, d-x, y)$.
As a consequence of this martingale property, if the producer already satisfies the relation (4.19) in the day-ahead market, and if the initial intraday price is the day-ahead price, her initial trading rate on the intraday market will be zero. And thus, on average, her trading rate will be zero.
The martingale property of the trading rate process is actually closely related to the martingale dynamics of the unaffected price $\hat{P}$ in (4.6). As we shall see in Section 4.4 where we consider jumps on price, making $\hat{P}$ a sub- or supermartingale, the optimal trading rate will inherit the converse sub- or supermartingale property. □

4.3.2 Approximate solution

We go back to the original execution problem with the non-negativity constraint on the production quantity. As pointed out above, there is no explicit solution in this case, due to the form of the terminal cost function C^+. The strategy is then to use the explicit control consisting in the trading rate $\hat{q}$ derived in (4.16), and of the truncated nonnegative production quantity:

$$\tilde{\xi}_T^* \;:=\; \hat{\xi}_T \mathbf{1}_{\hat{\xi}_T \geq 0} \;=\; \hat{\xi}^{tr+}(D_T^{t,d} - \hat{X}_T^{t,x,y,d}), \tag{4.21}$$

with $\hat{\xi}_T$ defined in (4.17) from the auxiliary problem. In other words, we follow the trading rate strategy $\hat{q}$ determined from the problem without constraint on the final production quantity, and at the terminal date use the production leverage if the final inventory $\hat{X}_T^{t,x,y,d}$ is below the terminal demand $D_T^{t,d}$, by choosing a quantity proportional to this spread $D_T^{t,d} - \hat{X}_T^{t,x,y,d}$. The aim of this section is to measure the relevance of this approximate strategy $(\hat{q}, \tilde{\xi}_T^*) \in \mathcal{A} \times L^0_+(\mathcal{F}_T)$ with respect to the optimal execution problem (4.9) by estimating the induced error:

$$\mathcal{E}_1(t,x,y,d) \;:=\; J(t,x,y,d;\hat{q},\tilde{\xi}_T^*) - v(t,x,y,d),$$

for $(t,x,y,d) \in [0,T] \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}$. We also measure the approximation error on the value function:

$$\mathcal{E}_2(t,x,y,d) \;:=\; v(t,x,y,d) - \tilde{v}(t,x,y,d).$$

Notice that if $\hat{\xi}_T \geq 0$ a.s., i.e. $D_T^{t,d} \geq \hat{X}_T^{t,x,y,d}$ a.s. (which is not true), and so $\tilde{\xi}_T^* = \hat{\xi}_T$, then clearly $(\hat{q}, \hat{\xi}_T)$ would be the solution to (4.9), and so $\mathcal{E}_1(t,x,y,d) = \mathcal{E}_2(t,x,y,d) = 0$. Actually, these errors depend on the probability of the event: $\{\hat{X}_T^{t,x,y,d} > D_T^{t,d}\}$, and we have the following estimate:

Proposition 4.2. *For all $(t,x,y,d) \in [0,T] \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}$, we have*

$$0 \;\leq\; \mathcal{E}_i(t,x,y,d) \;\leq\; \frac{\eta r(\eta,\beta)}{2\beta} V(T-t)\psi\Big(\frac{m(T-t,d-x,y)}{\sqrt{V(T-t)}}\Big), \quad i=1,2, \tag{4.22}$$

where

$$\psi(z) \;:=\; (z^2+1)\Phi(-z) - z\phi(z), \quad z \in \mathbb{R},$$

with $\phi = \Phi'$ the density of the standard normal distribution, and

$$m(t,d,y) \;:=\; \frac{(\nu t + 2\gamma)(\mu t + d) + yt}{(r(\eta,\beta)+\nu)t + 2\gamma}, \tag{4.23}$$

$$V(t) \;:=\; \int_0^t \frac{\sigma_0^2 s^2 + \sigma_d^2(\nu s + 2\gamma)^2 + 2\rho\sigma_0\sigma_d s(\nu s + 2\gamma)}{\big[(r(\eta,\beta)+\nu)s + 2\gamma\big]^2} ds \;\geq\; 0. \tag{4.24}$$

Proof. By definition of the value functions v and $\tilde{v}$, recalling that $(\hat{q}, \hat{\xi}_T)$ is an optimal control for $\tilde{v}$, and since $(\hat{q}, \tilde{\xi}_T^*) \in \mathcal{A} \times L^0_+(\mathcal{F}_T)$, we have:

$$J(t,x,y,d;\hat{q},\hat{\xi}_T) \;=\; \tilde{v}(t,x,y,d) \;\leq\; v(t,x,y,d) \;\;\leq\;\; J(t,x,y,d;\hat{q},\tilde{\xi}_T^*),$$

for all $(t,x,y,d) \in [0,T] \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}$. This clearly implies that both errors $\mathcal{E}_1$ and $\mathcal{E}_2$ are nonnegative, and

$$\max(\mathcal{E}_1(t,x,y,d), \mathcal{E}_2(t,x,y,d)) \;\;\leq\;\; \mathcal{E}(t,x,y,d) \;:=\; J(t,x,y,d;\hat{q},\tilde{\xi}_T^*) - J(t,x,y,d;\hat{q},\hat{\xi}_T).$$

We now focus on the upper bound for $\mathcal{E}$. By definition of J in (4.10), $\hat{\xi}_T$ and $\tilde{\xi}_T^*$ in (4.17) and (4.21), we have

$$\begin{aligned}
\mathcal{E}(t,x,y,d) \;&=\; \mathbb{E}\Big[C(D_T^{t,d} - \hat{X}_T^{t,x,y,d}, \tilde{\xi}_T^*) - C(D_T^{t,d} - \hat{X}_T^{t,x,y,d}, \hat{\xi}_T)\Big] \\
&=\; \mathbb{E}\Big[C(D_T^{t,d} - \hat{X}_T^{t,x,y,d}, \hat{\xi}^+(D_T^{t,d} - \hat{X}_T^{t,x,y,d})) \\
&\qquad\qquad - C(D_T^{t,d} - \hat{X}_T^{t,x,y,d}, \hat{\xi}(D_T^{t,d} - \hat{X}_T^{t,x,y,d}))\Big] \\
&=\; \mathbb{E}\Big[C^+(D_T^{t,d} - \hat{X}_T^{t,x,y,d}) - \tilde{C}(D_T^{t,d} - \hat{X}_T^{t,x,y,d})\Big] \\
&=\; \frac{\eta r(\eta,\beta)}{2\beta}\mathbb{E}\Big[\big(D_T^{t,d} - \hat{X}_T^{t,x,y,d}\big)^2 \mathbf{1}_{D_T^{t,d} - \hat{X}_T^{t,x,y,d} < 0}\Big],
\end{aligned} \tag{4.25}$$

from the definitions and expressions of C^+ and $\tilde{C}$ in (4.3), (4.13) and (4.15). Now, from (4.20) and by integration, we obtain the explicit (path-dependent) form of the optimal trading rate control:

$$\begin{aligned}
\hat{q}_s \;&=\; \hat{q}_t + \int_t^s \frac{r(\eta,\beta)\sigma_d}{(r(\eta,\beta)+\nu)(T-u)+2\gamma} dB_u \\
&\quad - \int_t^s \frac{\sigma_0}{(r(\eta,\beta)+\nu)(T-u)+2\gamma} dW_u, \quad t \leq s \leq T,
\end{aligned}$$

with $\hat{q}_t = \hat{q}(T-t, d-x, y)$. We then obtain the expression of the final spread between demand and inventory:

$$\begin{aligned}
D_T^{t,d} - \hat{X}_T^{t,x,y,d} \;&=\; d - x + \mu(T-t) + \int_t^T \sigma_d dB_s - \int_t^T \hat{q}_s ds \\
&=\; m(T-t, d-x, y) \;+ \int_t^T \frac{\sigma_d(\nu(T-s)+2\gamma)}{(r(\eta,\beta)+\nu)(T-s)+2\gamma} dB_s \\
&\quad + \int_t^T \frac{\sigma_0(T-s)}{(r(\eta,\beta)+\nu)(T-s)+2\gamma} dW_s,
\end{aligned}$$

by Fubini's theorem, and with

$$m(t,d,y) \;:=\; d + \mu t - t\hat{q}(t,d,y),$$

which is explicitly written as in (4.23) from the expression (4.16) of $\hat{q}$. Thus, $D_T^{t,d} - \hat{X}_T^{t,x,y,d}$ follows a normal distribution law with mean $m(T-t, d-x, y)$ and variance $V(T-t)$ given by (4.24), and from (4.25), we deduce that

$$\mathcal{E}(t,x,y,d) \;=\; \frac{\eta r(\eta,\beta)}{2\beta} V(T-t)\psi\Big(\frac{m(T-t,d-x,y)}{\sqrt{V(T-t)}}\Big),$$

while the probability that the final inventory is larger than the terminal demand is:

$$\mathbb{P}\big[D_T^{t,d} - \hat{X}_T^{t,x,y,d} < 0\big] \;=\; \Phi\Big(-\frac{m(T-t,d-x,y)}{\sqrt{V(T-t)}}\Big). \tag{4.26}$$

□

Error asymptotics. We now investigate the accuracy of the upper bound in (4.22)

$$\bar{\mathcal{E}}(T-t,d-x,y) \;:=\; \frac{\eta r(\eta,\beta)}{2\beta} V(T-t)\psi\Big(\frac{m(T-t,d-x,y)}{\sqrt{V(T-t)}}\Big).$$

It is well-known (see e.g. Section 14.8 in [75]) that

$$z\Phi(-z) \;\leq\; \phi(z), \quad \forall z \in \mathbb{R}, \tag{4.27}$$

from which we easily see that ψ is non-increasing, convex, and $\psi(\infty) = 0$. Thus, $\bar{\mathcal{E}}(T-t, d-x, y)$ decreases to zero for large $m(T-t, d-x; y)$ or small $V(T-t)$. We shall study its asymptotics in three limiting cases: (i) the time to maturity $T-t$ is small, (ii) the initial demand spread $d-x$ is large, (iii) the initial quoted price y is large. We prove that the error bound $\bar{\mathcal{E}}(T-t, d-x, y)$, and thus $\mathcal{E}_1(t,x,y,d)$, $\mathcal{E}_2(t,x,y,d)$, converge to zero at least with an exponential rate of convergence in these limiting regimes:

Proposition 4.3. *(i) For all $(x,y,d) \in \mathbb{R}\times\mathbb{R}\times\mathbb{R}$ with $d > x$, we have*

$$\limsup_{T-t\downarrow 0}(T-t)\ln\bar{\mathcal{E}}(T-t,d-x,y) \;\leq\; -\frac{1}{2}\Big(\frac{d-x}{\sigma_d}\Big)^2. \tag{4.28}$$

(ii) For all $(t,y) \in [0,T)\times\mathbb{R}$, we have

$$\limsup_{d-x\to\infty}\frac{1}{(d-x)^2}\ln\bar{\mathcal{E}}(T-t,d-x,y) \;\leq\; -\frac{1}{2}\frac{m_\infty^2(T-t)}{V(T-t)}, \tag{4.29}$$

where

$$m_\infty(t) \;=\; \frac{\nu t + 2\gamma}{\big(r(\eta,\beta)+\nu\big)t + 2\gamma}.$$

(iii) For all $(t,x,d) \in [0,T) \times \mathbb{R} \times \mathbb{R}$, we have

$$\limsup_{y\to\infty} \frac{1}{y^2} \ln \bar{\mathcal{E}}(T-t,d-x,y) \;\leq\; -\frac{1}{2}\frac{n_\infty^2(T-t)}{V(T-t)}, \tag{4.30}$$

where

$$n_\infty(t) \;=\; \frac{t}{(r(\eta,\beta)+\nu)t+2\gamma}.$$

Proof. From (4.27), we have:

$$0 \;\leq\; \psi(z) \;\leq\; z^{-1}\phi(z), \quad \forall z > 0.$$

Notice that in the three asymptotic regimes (i) (with $d-x>0$), (ii), and (iii), the quantity $m(T-t,d-x,y)$ is positive, and we thus have:

$$\bar{\mathcal{E}}(T-t,d-x,y) \;\leq\; \frac{\eta r(\eta,\beta)}{2\beta}\frac{V(T-t)^{\frac{3}{2}}}{m(T-t,d-x,y)}\phi\Big(\frac{m(T-t,d-x,y)}{\sqrt{V(T-t)}}\Big). \tag{4.31}$$

(i) For small time to maturity $T-t$, we see that $m(T-t,d-x,y)$ converges to $d-x>0$, while $V(T-t) \sim \sigma_d^2(T-t)$, i.e. $V(T-t)/\sigma_d^2(T-t)$ converges to 1. This shows from (4.31) that, when $T-t$ goes to zero, the error bound $\bar{\mathcal{E}}(T-t,d-x,y)$, converges to zero at least with an exponential rate of convergence, namely the one given by (4.28).
(ii) For large demand spread $d-x$, we see that $m(T-t,d-x,y) \sim m_\infty(T-t)(d-x)$, i.e. the ratio $m(T-t,d-x,y)/m_\infty(T-t)(d-x)$ converges to 1 when $d-x$ goes to infinity. This shows from (4.31) that, when $d-x$ goes to infinity, the error bound $\bar{\mathcal{E}}(T-t,d-x,y)$, converges to zero at least with an exponential rate of convergence, namely the one given by (4.29).
(iii) For large y, we see that $m(T-t,d-x,y) \sim n_\infty(T-t)y$, i.e. the ratio $m(T-t,d-x,y)/n_\infty(T-t)y$ converges to 1 when y goes to infinity. This shows from (4.31) that, when $d-x$ goes to infinity, the error bound $\bar{\mathcal{E}}(T-t,d-x,y)$ converges to zero at least with an exponential rate of convergence, namely the one given by (4.30). □

Interpretation. Recall from (4.26) that

$$\mathbb{P}\big[D_T^{t,d} < \hat{X}_T^{t,x,y,d}\big] \;=\; \Phi\Big(-\frac{m(T-t,d-x,y)}{\sqrt{V(T-t)}}\Big),$$

and thus following the same arguments as in the above proof, we have:

(i)

$$\limsup_{T-t\downarrow 0}(T-t)\ln\mathbb{P}\big[D_T^{t,d} < \hat{X}_T^{t,x,y,d}\big] \;=\; -\frac{1}{2}\Big(\frac{d-x}{\sigma_d}\Big)^2, \tag{4.32}$$

for all $(x,y,d) \in \mathbb{R}\times\mathbb{R}\times\mathbb{R}$ with $d>x$. We observe that the rate in the rhs of (4.28) (or (4.32)) depends only on the demand volatility σ_d and the initial demand spread

$d - x$. Moreover, it is all the larger, the smaller σ_d is, and the larger $d - x$ is. This means that the terminal demand will stay with very high probability above the final inventory once we are near from the maturity with a low volatile demand, initially larger than the inventory, in which case, the explicit strategy $(\hat{q}, \tilde{\xi}_T^*)$ approximates very accurately the optimal strategy (q^*, ξ_T^*).

(ii)

$$\limsup_{d-x\to\infty} \frac{1}{(d-x)^2} \ln \mathbb{P}\big[D_T^{t,d} < \hat{X}_T^{t,x,y,d}\big] \quad = \quad -\frac{1}{2}\frac{m_\infty^2(T-t)}{V(T-t)}, \tag{4.33}$$

for all $(t, y) \in [0, T) \times \mathbb{R}$, The rate in the rhs of (4.29) (or (4.33)) is all the larger, the smaller the volatilities σ_0 and σ_d of the electricity price and demand are. Again, we have the same interpretation than in the asymptotic regime (i), and this means that the explicit strategy $(\hat{q}, \tilde{\xi}_T^*)$ approximates very accurately the optimal strategy (q^*, ξ_T^*) in the limiting regime when the initial demand spread is large, and the volatilities are small.

(iii)

$$\limsup_{y\to\infty} \frac{1}{y^2} \ln \mathbb{P}\big[D_T^{t,d} < \hat{X}_T^{t,x,y,d}\big] \quad = \quad -\frac{1}{2}\frac{n_\infty^2(T-t)}{V(T-t)}, \tag{4.34}$$

for all $(t, x, d) \in [0, T) \times \mathbb{R} \times \mathbb{R}$. In the limiting regime where the initial quoted price y is large, the agent has a strong incentive to sell energy on the intraday market, which leads to a final inventory staying under the final demand with high probability, and thus to a very accurate approximate strategy $(\hat{q}, \tilde{\xi}_T^*)$. As in case (ii), this accuracy is strengthened for small volatilities σ_0 and σ_d of the electricity price and demand. □

4.3.3 Numerical results

Numerical tests

We measure quantitatively the accuracy of the error bound derived in the previous paragraph with some numerical tests. Let us fix the following parameter values: $\sigma_0 = 1/60$ €·(MW)$^{-1}$ · $s^{-1/2}$, $\sigma_d = 1000/60$ MW·$s^{-1/2}$, $\beta = 0.002$ €·(MW)$^{-2}$, $\eta = 200$ €·(MW)$^{-2}$, $\mu = 0$ MW·s^{-1}, $\nu = 10^{-10}$€·(MW)$^{-2}$, $\gamma = 10^{-10}$€·s·(MW)$^{-2}$ and $\rho = 0.8$.
We start from the initial time $t = 0$, with a zero inventory $X_0 = 0$, and vary respectively the maturity T, the initial demand D_0 and the initial price Y_0. We compute the probability for the final inventory to exceed the final demand $\mathbb{P}[\hat{X}_T > D_T]$, the approximate value function $\tilde{v}(0, X_0, Y_0, D_0)$, and the error bound $\bar{\mathcal{E}}(T, D_0 - X_0, Y_0)$. The results are reported in Table 4.1 when varying T, in Table 4.2 when varying D_0 and in Table 4.3 when varying Y_0.
Table 4.1 shows that for time to maturity less than $T = 24$h, the probability for the final inventory to exceed the final demand is very small, and consequently the error bound is rather negligible. When the time horizon increases, the agent has the possibility to spread over time her trading strategies for reducing the price impact, and purchase more energy, in which case the probability for the final inventory to exceed the demand increases.

T (h)	$\mathbb{P}[\hat{X}_T > D_T]$	$\tilde{v}(0, X_0, Y_0, D_0)$ (€)	$\bar{\mathcal{E}}(T, D_0 - X_0, Y_0)$ (€)
1	$< 10^{-16}$	1.88×10^6	$< 10^{-16}$
8	$< 10^{-16}$	1.88×10^6	$< 10^{-16}$
24	$< 10^{-16}$	1.89×10^6	4.16×10^{-12}
50	7.72×10^{-13}	1.90×10^6	2.48×10^{-4}

Table 4.1 – $Y_0 = 50$ €·(MW)$^{-1}$ and $D_0 = 50,000$ MW

D_0 (MW)	$\mathbb{P}[\hat{X}_T > D_T]$	$\tilde{v}(0, X_0, Y_0, D_0)$ (€)	$\bar{\mathcal{E}}(T, D_0 - X_0, Y_0)$ (€)
500	$< 10^{-16}$	-5.86×10^5	4.16×10^{-12}
$5,000$	$< 10^{-16}$	-3.62×10^5	4.16×10^{-12}
$50,000$	$< 10^{-16}$	1.89×10^6	4.16×10^{-12}
$500,000$	$< 10^{-16}$	2.44×10^7	4.16×10^{-12}

Table 4.2 – $T = 24$ h and $Y_0 = 50$ €·(MW)$^{-1}$

Table 4.2 shows that the probability for the final inventory to exceed the final demand and the error bound are not much sensitive to the variations of the initial positive demand D_0. Actually, the main impact is caused by the initial stock price, as observed in Table 4.3.

Y_0 (€·(MW)$^{-1}$)	$\mathbb{P}[\hat{X}_T > D_T]$	$\tilde{v}(0, X_0, Y_0, D_0)$(€)	$\bar{\mathcal{E}}(T, D_0 - X_0, Y_0)$ (€)
500	$< 10^{-16}$	2.51×10^6	$< 10^{-16}$
50	$< 10^{-16}$	1.89×10^6	4.16×10^{-12}
40	9.51×10^{-15}	1.61×10^6	3.80×10^{-4}
30	4.57×10^{-10}	1.29×10^6	1.30×10^{-2}
20	2.23×10^{-5}	9.13×10^5	1.26×10^3

Table 4.3 – $T = 24$ h and $D_0 = 50,000$ MW

For small initial electricity price Y_0, the agent will buy more energy in the intraday market and produce less. Therefore, the inventory will overtake with higher probability the demand, in which case the approximate value function can be significantly different from the original one, as observed from the error bound in Table 4.3 for $Y_0 = 20$.

Simulations

We plot trajectories of some relevant quantities that we simulate with the following set of parameters: $\sigma_0 = 1/60$ €·(MW)$^{-1}$·$s^{-1/2}$, $\sigma_d = 1000/60$ MW·$s^{-1/2}$, $\beta = 0.002$ €·(MW)$^{-2}$, $\eta = 100$ €·(MW)$^{-2}$, $\mu = 0$ MW·s^{-1}, $\rho = 0.8$, $\nu = 4.00 \times 10^{-5}$€·(MW)$^{-2}$, $\gamma = 2.22$€·s·(MW)$^{-2}$, $T = 24$h, $X_0 = 0$, $D_0 = 50,000$ MW and $Y_0 = 50$ €·(MW)$^{-1}$.
For such parameter values, the probability $\mathbb{P}[\hat{X}_T > D_T]$ is bounded above by 10^{-16}, the error $\bar{\mathcal{E}}(0, D_0 - X_0, Y_0)$ is bounded by 2.82×10^{-10}€, and

$$\tilde{v}(0, X_0, Y_0, D_0) \quad = \quad 1916700€.$$

The executed strategy $(\hat{q}, \hat{\xi}_T^*)$ can then be considered as very close to the optimal strategy.

Figure 4.1 represents the evolution of the trading rate control $(\hat{q}_t)_{t\in[0,T]}$ derived in (4.16) for a given trajectory of price and demand, and this is consistent with the martingale property as shown in Proposition 4.1. Figure 4.2 represents a simulation of the quoted price $\hat{Y}_t$ with impact and of the unaffected price $\hat{P}_t$. Due to the buying strategy, i.e. positive $\hat{q}$, we observe that the quoted price $\hat{Y}$ is larger than $\hat{P}$. In Figure 4.3, we plot the evolution of the optimal inventory $(\hat{X}_t)_{t\in[0,T)}$, and of the forecast residual demand $(D_t)_{t\in[0,T]}$. We see that $\hat{X}_t$ is increasing, with a growth rate which looks constant as pointed out in Remark 4.4. At final time, if $\hat{X}_T < D_T$ (which is the case in our simulation), the agent uses her production leverage $\hat{\xi}_T$, and achieves a final inventory: $\hat{X}_T + \hat{\xi}_T$, which is represented by the peak at time T. From the expression (4.17) of $\hat{\xi}_T$, the final imbalance cost is equal to

$$D_T - \hat{X}_T - \hat{\xi}_T \quad = \quad \frac{\beta}{\eta + \beta}(D_T - \hat{X}_T),$$

and is then positive, as shown in Figure 4.3.

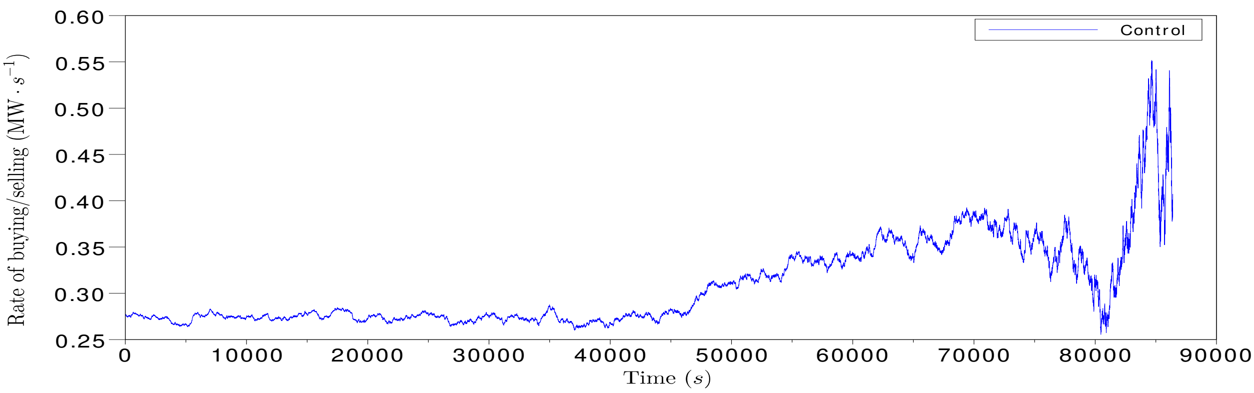

Figure 4.1 – Evolution of the trading rate control $\hat{q}$

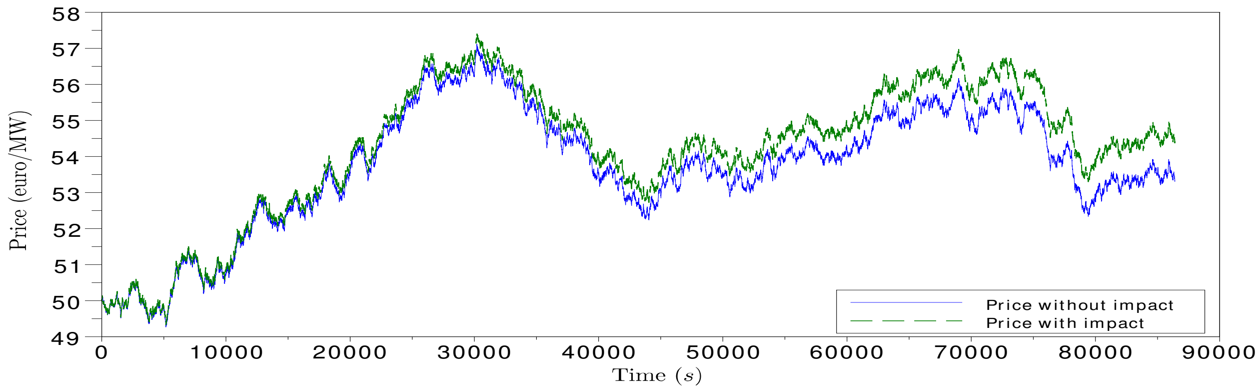

Figure 4.2 – Simulation of the quoted impacted price $\hat{Y}$ and of the unaffected price $\hat{P}$.

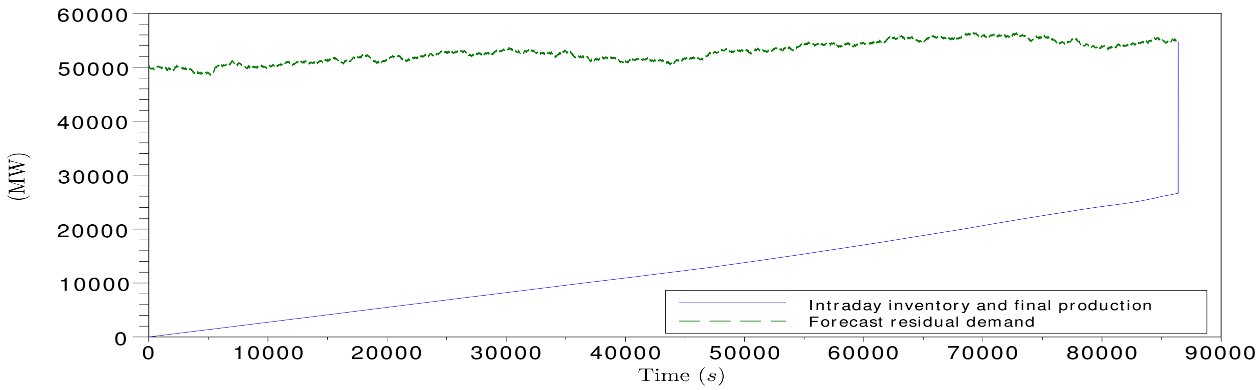

Figure 4.3 – Evolution of the inventory $\hat{X}$ and of the forecast residual demand D.

4.4 Jumps in the residual demand forecast

In this section, we incorporate the case where the residual demand forecast is subject to sudden changes induced by prediction errors on renewable production, which may be quite large. Our aim is to study the impact on the strategies obtained in the previous section, and we shall also neglect the delay in thermal plants production.

The sudden changes in the demand forecast are modeled via a compound Poisson process $N_t = (N_t^+, N_t^-)_{t\geq 0}$ with intensity $\lambda > 0$, where N_t^+ is the counting process associated to positive jumps of the demand forecast with size $\delta^+ > 0$, occurring with probability $p^+ \in [0,1]$, while N_t^- is the counting process associated to negative jumps of the demand forecast with size $\delta^- < 0$, occurring with probability $p^- = 1 - p^+$. We denote by $\delta := \delta^+ p^+ + \delta^- p^-$ the mean of the jump size of the demand forecast. The dynamics of the residual demand forecast D is then given by:

$$dD_t \;=\; \mu dt + \sigma_d dB_t + \delta^+ dN_t^+ + \delta^- dN_t^-, \tag{4.35}$$

where we add a jump component with respect to the model in (4.8). Moreover, as soon as a jump in the residual demand forecast occurs, this is impacted into the intraday electricity price since the main producers are assumed to have access to the whole updated forecast. We thus model the unaffected electricity price by:

$$\hat{P}_t \;=\; \hat{P}_0 + \sigma_0 W_t + \pi^+ N_t^+ + \pi^- N_t^-, \tag{4.36}$$

where we add with respect to the Bachelier model in (4.6) a jump component of size $\pi^+ > 0$ (resp. $\pi^- < 0$) when the jump on residual demand is positive (resp. negative), which means that a higher (resp. lower) demand induces an increase (resp. drop) of price. We denote by $\pi := \pi^+ p^+ + \pi^- p^-$ the mean of the jump size of the intraday price. Given a trading rate $q \in \mathcal{A}$, the dynamics of the quoted price Y is then governed by

$$dY_t \;=\; \nu q_t dt + \sigma_0 dW_t + \pi^+ dN_t^+ + \pi^- dN_t^-. \tag{4.37}$$

By considering this simplified modeling of demand forecast subject to sudden shift in terms of a Poisson process, we do not have additional state variables with respect to the no jump case of the previous section. Let us then denote by $v = v^{(\lambda)}(t,x,y,d)$ the value function to the optimal execution problem (4.4) with cost functional $J = J^{(\lambda)}(t,x,y,d,q,\xi)$, where we stress the dependence in λ for taking into account jumps in demand forecast. The value function in the no jump case derived in the previous section is denoted by $v = v^{(0)}$.

As in the case with no jumps, there is no explicit solution to $v^{(\lambda)}$ due to the non-negativity constraint on the final production: we shall first study the auxiliary execution problem without sign constraint on the final production, then provide an approximate solution to the original one with an estimation of the induced error approximation, and with some numerical illustrations. We compare the results with the no jump case by focusing on the impact of the jump components.

4.4.1 Auxiliary optimal execution problem

Similarly as in Subsection 4.3.1, we consider the optimal execution problem without non-negativity constraint on the final production, denoted by $\tilde{v} = \tilde{v}^{(\lambda)}(t,x,y,d)$.
As in Theorem 4.1 for the case of the value function $\tilde{v}^{(0)}$ without jumps, we have an explicit solution to this auxiliary problem.

Theorem 4.2. *The value function to the auxiliary optimization problem is explicitly given by:*

$$
\begin{aligned}
& \tilde{v}^{(\lambda)}(t,x,y,d) \\
= \;& \tilde{v}^{(0)}(t,x,y,d) \\
& + \frac{\lambda}{2} \frac{r(\eta,\beta)(T-t)\big(\pi(T-t) + 2\delta(\nu(T-t)+2\gamma)\big)}{\big(r(\eta,\beta)+\nu\big)(T-t)+2\gamma}(d-x) \\
& - \frac{\lambda}{2} \frac{(T-t)^2\big(\pi - 2r(\eta,\beta)\delta\big)}{\big(r(\eta,\beta)+\nu\big)(T-t)+2\gamma} y \\
& + \lambda\gamma \frac{p^+(\pi^+ - r(\eta,\beta)\delta^+)^2 + p^-(\pi^- - r(\eta,\beta)\delta^-)^2}{\big(r(\eta,\beta)+\nu\big)^2} \ln\Big(1 + \frac{(r(\eta,\beta)+\nu)(T-t)}{2\gamma}\Big) \\
& - \frac{\lambda}{2} \frac{p^+((\pi^+)^2 - r(\eta,\beta)\delta^+(2\pi^+ + \nu\delta^+)) + p^-((\pi^-)^2 - r(\eta,\beta)\delta^-(2\pi^- + \nu\delta^-))}{r(\eta,\beta)+\nu}(T-t) \\
& + \frac{\lambda r(\eta,\beta)}{2} \frac{2\nu\mu\delta + \lambda((p^+)^2\delta^+(\pi^+ + \nu\delta^+) + (p^-)^2\delta^-(\pi^- + \nu\delta^-))}{r(\eta,\beta)+\nu}(T-t)^2 \\
& + \lambda^2\gamma r(\eta,\beta) \frac{r(\eta,\beta)\delta^2 + 2\nu p^+p^-\delta^+\delta^- - ((p^+)^2\delta^+\pi^+ + (p^-)^2\delta^-\pi^-)}{(r(\eta,\beta)+\nu)\big((r(\eta,\beta)+\nu)(T-t)+2\gamma\big)}(T-t)^2 \\
& + \frac{2\lambda\gamma r^2(\eta,\beta)\mu\delta}{(r(\eta,\beta)+\nu)\big((r(\eta,\beta)+\nu)(T-t)+2\gamma\big)}(T-t)^2 - \frac{\lambda^2\pi^2}{48\gamma}(T-t)^3 \\
& + \frac{\lambda^2 p^+p^- r(\eta,\beta)}{2} \frac{2\nu\delta^+\delta^- + \delta^-\pi^+ + \delta^+\pi^-}{(r(\eta,\beta)+\nu)(T-t)+2\gamma}(T-t)^3 \\
& + \frac{1}{8} \frac{4r(\eta,\beta)\mu\lambda\pi - \lambda^2\pi^2}{(r(\eta,\beta)+\nu)(T-t)+2\gamma}(T-t)^3,
\end{aligned}
$$

for $(t,x,y,d) \in [0,T]\times\mathbb{R}\times\mathbb{R}\times\mathbb{R}$, *with an optimal trading rate given in feedback form by:*

$$
\begin{aligned}
\hat{q}^{(\lambda)}_s &= \hat{q}^{(\lambda)}(T-s, D^{t,d}_s - \hat{X}^{t,x,y,d}_s, \hat{Y}^{t,x,y,d}_s), \quad t \le s \le T \\
\hat{q}^{(\lambda)}(t,d,y) &:= \hat{q}^{(0)}(t,d,y) + \lambda \frac{r(\eta,\beta)\delta t + \frac{\pi}{4\gamma}(r(\eta,\beta)+\nu)t^2}{(r(\eta,\beta)+\nu)t + 2\gamma} \\
&= \hat{q}^{(0)}(t, d+\lambda\delta t, y + \frac{\lambda}{2}\pi t) + \frac{\lambda\pi}{4\gamma}t,
\end{aligned} \tag{4.38}
$$

where $\hat{q}^{(0)}$ *is the optimal trading rate given in (4.16) in the case with no jump in the demand forecast. Here* $(\hat{X}^{t,x,y,d}, \hat{Y}^{t,x,y,d}, D^{t,d})$ *denotes the solution to (4.1)-(4.37)-(4.35) when using* *the feedback control $\hat{q}^{(\lambda)}$, and starting from (x, y, d) at time t. Finally, the optimal production quantity is given by:*

$$\hat{\xi}_T^{(\lambda)} \;=\; \frac{\eta}{\eta+\beta}\big(D_T^{t,d} - \hat{X}_T^{t,x,y,d}\big). \tag{4.39}$$

Proof. See Appendix. □

Interpretation. The expression of the optimal trading rate $\hat{q}_s^{(\lambda)}$, $s \in [t, T]$, as

$$\hat{q}_s^{(\lambda)} \;=\; \hat{q}_s^{(0)} + \lambda \frac{r(\eta,\beta)\delta(T-s) + \frac{\pi}{4\gamma}(r(\eta,\beta)+\nu)(T-s)^2}{(r(\eta,\beta)+\nu)(T-s)+2\gamma},$$

where $\hat{q}_s^{(0)} = \hat{q}^{(0)}(T-s, D_s^{t,d} - \hat{X}_s^{t,x,y,d}, \hat{Y}_s^{t,x,y,d})$ represents the optimal trading rate that the agent would use if she believes that the demand forecast will not jump, shows that under the information knowledge about jumps, the agent will purchase more (resp. less) electricity shares and this impact is all the larger, the larger the intensity λ of jumps, and the positive (resp. negative) mean δ and π of jump size in demand forecast and price are. On the other hand, the expression of $\hat{q}_s^{(\lambda)}$ as the sum of two terms:

$$\hat{q}_s^{(\lambda)} \;=\; \hat{q}^{(0)}\big(T-s, D_s^{t,d} + \lambda\delta(T-s), \hat{Y}_s^{t,x,y,d} + \frac{\lambda}{2}\pi(T-s)\big) \;+\; \frac{\lambda\pi}{4\gamma}(T-s), \tag{4.40}$$

can be interpreted as follows. The first term is analog to the optimal trading rate in the no jump case, with an adjustment $\lambda\delta(T-s)$ in the demand, which represents the expectation of the demand jump size up to the final horizon, and an adjustment $\frac{\lambda}{2}\pi(T-s)$ on the price, which represents half of the expectation of the price jump size up to the final horizon. The second term, $\frac{\lambda\pi}{4\gamma}(T-s)$, is deterministic, and linear in time, and we shall see on the simulations for some parameter values that it can be dominant with respect to the first stochastic term. Moreover, as in (4.19), we can write an equilibrium relation which indicates that this control aims at making the forecast intraday price and the forecast cost of production equal, in particular at terminal date T :

$$\hat{Y}_T^{t,x,y,d} + 2\gamma\hat{q}_T^{(\lambda)} \;=\; c'(\hat{\xi}_T^{(\lambda)}). \tag{4.41}$$

□

The unaffected price $\hat{P}$ in (4.36) is no more a martingale in presence of jumps, except when $\pi = 0$. It is actually a supermartingale when $\pi < 0$ (predominant negative jumps), and submartingale when $\pi > 0$ (predominant positive jumps). The next result shows that the optimal trading rate inherits the converse submartingale or supermartingale property of the price process.

Proposition 4.4. *The optimal trading rate process $(\hat{q}_s^{(\lambda)})_{t\le s\le T}$ in (4.38) is a supermartingale if $\pi > 0$, and a submartingale if $\pi < 0$. More precisely, the process $\{\hat{q}_s^{(\lambda)} + \frac{\lambda\pi}{2\gamma}(s-t), t \le s \le T\}$ is a martingale.*

Proof. Notice that $N_t^{\pm}$ is a Poisson process with intensity $\lambda p^{\pm}$, and let us introduce the compensated martingale Poisson process $\tilde{N}_t^{\pm} = N_t - \lambda p^{\pm} t$. By applying Itô's formula to the trading rate process $\hat{q}_s^{(\lambda)} = \hat{q}^{(\lambda)}(T-s, D_s^{t,d} - \hat{X}_s^{t,x,y,d}, \hat{Y}_s^{t,x,y,d})$, $t \leq s \leq T$, and from the dynamics (4.1), (4.35) and (4.37), we have:

$$
\begin{aligned}
d\hat{q}_s^{(\lambda)} \;=\; & \Big[-\frac{\partial \hat{q}^{(\lambda)}}{\partial t} + (\mu - \hat{q}^{(\lambda)})\frac{\partial \hat{q}^{(\lambda)}}{\partial d} + \nu \hat{q}^{(\lambda)} \frac{\partial \hat{q}^{(\lambda)}}{\partial y} \\
& \quad + \lambda p^{+}\big(\hat{q}^{(\lambda)}(.,. + \delta^{+}, . + \pi^{+}) - \hat{q}^{(\lambda)}\big) \\
& \quad + \lambda p^{-}\big(\hat{q}^{(\lambda)}(.,. + \delta^{-}, . + \pi^{-}) - \hat{q}^{(\lambda)}\big)\Big](T-s, D_s^{t,d} - \hat{X}_s^{t,x,y,d}, \hat{Y}_s^{t,x,y,d})ds \\
& + \frac{\partial \hat{q}^{(\lambda)}}{\partial d}(T-s, D_s^{t,d} - \hat{X}_s^{t,x,y,d}, \hat{Y}_s^{t,x,y,d})\sigma_d dB_s \\
& + \frac{\partial \hat{q}^{(\lambda)}}{\partial y}(T-s, D_s^{t,d} - \hat{X}_s^{t,x,y,d}, \hat{Y}_s^{t,x,y,d})\sigma_0 dW_s \\
& + \big[\hat{q}^{(\lambda)}(T-s, D_{s^-}^{t,d} + \delta^{+} - \hat{X}_s^{t,x,y,d}, \hat{Y}_{s^-}^{t,x,y,d} + \pi^{+}) \\
& \qquad\qquad - \hat{q}^{(\lambda)}(T-s, D_{s^-}^{t,d} - \hat{X}_s^{t,x,y,d}, \hat{Y}_{s^-}^{t,x,y,d})\big]d\tilde{N}_s^{+} \\
& + \big[\hat{q}^{(\lambda)}(T-s, D_{s^-}^{t,d} + \delta^{-} - \hat{X}_s^{t,x,y,d}, \hat{Y}_{s^-}^{t,x,y,d} + \pi^{-}) \\
& \qquad\qquad - \hat{q}^{(\lambda)}(T-s, D_{s^-}^{t,d} - \hat{X}_s^{t,x,y,d}, \hat{Y}_{s^-}^{t,x,y,d})\big]d\tilde{N}_s^{-}.
\end{aligned}
$$

Now, from the expression (4.38) of $\hat{q}^{(\lambda)}(t,d,y)$, we see that:

$$
\begin{aligned}
& -\frac{\partial \hat{q}^{(\lambda)}}{\partial t} + (\mu - \hat{q}^{(\lambda)})\frac{\partial \hat{q}^{(\lambda)}}{\partial d} + \nu \hat{q}^{(\lambda)} \frac{\partial \hat{q}^{(\lambda)}}{\partial y} \\
+\lambda\big(p^{+}\hat{q}^{(\lambda)}(.,. + \delta^{+}, . + \pi^{+}) + p^{-}\hat{q}^{(\lambda)}(.,. + \delta^{-}, . + \pi^{-}) - \hat{q}^{(\lambda)}\big) \;=\; & -\frac{\lambda \pi}{2\gamma},
\end{aligned}
$$

and then:

$$
\begin{aligned}
d\hat{q}_s^{(\lambda)} \;=\; & -\frac{\lambda \pi}{2\gamma} ds \\
& + \frac{r(\eta,\beta)\sigma_d}{(r(\eta,\beta)+\nu)(T-s)+2\gamma} dB_s - \frac{\sigma_0}{(r(\eta,\beta)+\nu)(T-s)+2\gamma} dW_s \\
& + \frac{r(\eta,\beta)\delta^{+} - \pi^{+}}{(r(\eta,\beta)+\nu)(T-s)+2\gamma} d\tilde{N}_s^{+} + \frac{r(\eta,\beta)\delta^{-} - \pi^{-}}{(r(\eta,\beta)+\nu)(T-s)+2\gamma} d\tilde{N}_s^{-}.
\end{aligned} \tag{4.42}
$$

This proves the required assertions of the proposition. □

Remark 4.5. The above supermartingale (or submartingale) property implies in particular that the mean of the optimal trading rate process $(\hat{q}_s^{(\lambda)})_{0 \leq s \leq T}$ is decreasing (or increasing) in time, and so that the trajectory of the optimal inventory mean $\mathbb{E}[\hat{X}_s^{0,x,y,d}]$, $0 \leq s \leq T$, is concave (or convex). Moreover, from the martingale property of $\hat{q}_s^{(\lambda)} + \frac{\lambda \pi}{2\gamma} s$, $0 \leq s \leq T$, we have: $\mathbb{E}[\hat{q}_s^{(\lambda)}] = \hat{q}^{(\lambda)}(T, d-x, y) - \frac{\lambda\pi}{2\gamma}s$ for $0 \leq s \leq T$. Fix d, x, y, and let us then denote by $\bar{s}^{(\lambda)} := \frac{2\gamma}{\lambda\pi}\hat{q}^{(\lambda)}(T, d-x, y)$, which is explicitly written as:

$$\bar{s}^{(\lambda)} = \frac{T}{2} + \frac{1}{\lambda\pi}\frac{\big(r(\eta,\beta)\mu + \lambda(r(\eta,\beta)\delta - \frac{\pi}{2})\big)T + r(\eta,\beta)(d-x) - y}{1 + \frac{(r(\eta,\beta)+\nu)T}{2\gamma}}$$

We have the following cases:

— $\bar{s}^{(\lambda)} \leq 0$ and $\pi > 0$: this may arise for large y, or $d << x$, or $r(\eta,\beta)\delta << \pi/2$. In this extreme case, $\frac{d\mathbb{E}[\hat{X}_s^{0,x,y,d}]}{ds} = \mathbb{E}[\hat{q}_s^{(\lambda)}] \leq 0$ for $0 \leq s \leq T$, i.e. the trajectory of $\mathbb{E}[\hat{X}_s^{0,x,y,d}]$, $0 \leq s \leq T$, is decreasing, which means that the agent will "always" sell electricity shares since she takes advantage of high price, in order to decrease her inventory for approaching the demand, and because in average, the jump size of the demand is much lower than the positive jump size of the price.

— $\bar{s}^{(\lambda)} \leq 0$ and $\pi < 0$: this may arise for small y, or $d >> x$, or $r(\eta,\beta)\delta >> \pi/2$. In this extreme case, $\frac{d\mathbb{E}[\hat{X}_s^{0,x,y,d}]}{ds} = \mathbb{E}[\hat{q}_s^{(\lambda)}] \geq 0$ for $0 \leq s \leq T$, i.e. the trajectory of $\mathbb{E}[\hat{X}_s^{0,x,y,d}]$, $0 \leq s \leq T$, is increasing, which means that the agent will "always" buy electricity shares since she takes advantage of low price, in order to increase her inventory for approaching the demand, and because in average, the jump size of the price is much lower than the jump size of the demand.

— $\bar{s}^{(\lambda)} \geq T$ and $\pi > 0$: this may arise for $r(\eta,\beta)\delta >> \pi/2$, $d >> x$ or small y. In this other extreme case, the trajectory of $\mathbb{E}[\hat{X}_s^{0,x,y,d}]$, $0 \leq s \leq T$, is increasing, which means that the agent will "always" buy electricity shares at low price in order to approach the residual demand at final time.

— $\bar{s}^{(\lambda)} \geq T$ and $\pi < 0$: this may arise for $r(\eta,\beta)\delta << \pi/2$, $d << x$ or large y. The trajectory of $\mathbb{E}[\hat{X}_s^{0,x,y,d}]$, $0 \leq s \leq T$, is decreasing, which means that the agent will "always" sell electricity shares at high price in order to approach the residual demand at final time.

— $0 < \bar{s}^{(\lambda)} < T$: in this regular case, it is interesting to comment on the two subcases:
  — if $\pi > 0$, the trajectory of $s \mapsto \mathbb{E}[\hat{X}_s^{0,x,y,d}]$ is increasing for $s \leq \bar{s}^{(\lambda)}$ and then decreasing for $\bar{s}^{(\lambda)} < s \leq T$. This means that the agent starts by purchasing electricity shares for taking profit of the positive price jumps (which have more impact than the negative price jumps as $p^+\pi^+ + p^-\pi^- > 0$), and then resells shares in order to achieve the equilibrium relation (4.41).
  — if $\pi < 0$, i.e. the negative jumps have more impact than the positive ones: the agent starts by selling electricity shares and then purchases shares.

□

4.4.2 Approximate solution

We turn back to the original optimal execution problem with the non-negativity constraint on the final production, and as in Section 4.3.2, we use the approximate strategy consisting in the trading rate $\hat{q}^{(\lambda)}$ derived in (4.38), and of the truncated nonnegative final production:

$$\tilde{\xi}_T^{(\lambda),*} \;:=\; \hat{\xi}_T^{(\lambda)}\mathbf{1}_{\hat{\xi}_T^{(\lambda)}\geq 0} \;=\; \hat{\xi}^{tr+}(D_T^{t,d} - \hat{X}_T^{t,x,y,d}),$$

with $\hat{\xi}_T^{(\lambda)}$ given in (4.39). We measure the relevance of this strategy $(\hat{q}^{(\lambda)}, \tilde{\xi}_T^{(\lambda),*}) \in \mathcal{A} \times L^0_+(\mathcal{F}_T)$ by estimating the induced error:

$$\mathcal{E}_1^{(\lambda)}(t,x,y,d) \;:=\; J^{(\lambda)}(t,x,y,d;\hat{q}^{(\lambda)},\tilde{\xi}_T^{(\lambda),*}) - v^{(\lambda)}(t,x,y,d),$$

for $(t,x,y,d) \in [0,T]\times\mathbb{R}\times\mathbb{R}\times\mathbb{R}$, and also measure the approximation error on the value functions:

$$\mathcal{E}_2^{(\lambda)}(t,x,y,d) \;:=\; v^{(\lambda)}(t,x,y,d) - \tilde{v}^{(\lambda)}(t,x,y,d).$$

Proposition 4.5. *For all $(t,x,y,d) \in [0,T]\times\mathbb{R}\times\mathbb{R}\times\mathbb{R}$, we have*

$$0 \;\leq\; \mathcal{E}_i^{(\lambda)}(t,x,y,d) \;\leq\; \frac{\eta r(\eta,\beta)}{2\beta}V(T-t)\mathbb{E}\Big[\psi\Big(\frac{m^{(\lambda)}(T-t,d-x,y)+\Sigma_T^{-,t}}{\sqrt{V(T-t)}}\Big)\Big] \tag{4.43}$$

for $i=1,2$, where ψ, m, V are defined in Proposition 4.2,

$$\begin{aligned} m^{(\lambda)}(t,d,y) \;&=\; m\Big(t,d,y+\lambda\big(\frac{\pi}{2}-r(\eta,\beta)\delta\big)t\Big) \\ &\quad + \lambda\frac{r(\eta,\beta)\delta-\pi}{r(\eta,\beta)+\nu}\Big[t - \frac{2\gamma}{r(\eta,\beta)+\nu}\ln\Big(1+\frac{r(\eta,\beta)+\nu}{2\gamma}t\Big)\Big], \end{aligned} \tag{4.44}$$

and

$$\Sigma_T^{-,t} \;=\; \int_t^T \frac{\delta^-(\nu(T-s)+2\gamma)+\pi^-(T-s)}{(r(\eta,\beta)+\nu)(T-s)+2\gamma}dN_s^- \;\leq\; 0, \quad a.s.$$

Proof. By the same arguments as in Proposition 4.2, we have

$$0 \leq \;\mathcal{E}_i^{(\lambda)}(t,x,y,d) \;\leq\; \mathcal{E}^{(\lambda)}(t,x,y,d) \;:=\; J^{(\lambda)}(t,x,y,d;\hat{q}^{(\lambda)},\tilde{\xi}_T^{(\lambda),*}) - J^{(\lambda)}(t,x,y,d;\hat{q}^{(\lambda)},\hat{\xi}_T^{(\lambda)}),$$

for $i=1,2$, and

$$\mathcal{E}^{(\lambda)}(t,x,y,d) \;=\; \frac{\eta r(\eta,\beta)}{2\beta}\mathbb{E}\Big[\big(D_T^{t,d}-\hat{X}_T^{t,x,y,d}\big)^2\mathbf{1}_{D_T^{t,d}-\hat{X}_T^{t,x,y,d}<0}\Big], \tag{4.45}$$

for $(t,x,y,d) \in [0,T]\times\mathbb{R}\times\mathbb{R}\times\mathbb{R}$. Now, recall from (4.42) that:

$$\begin{aligned} d\hat{q}_s^{(\lambda)} \;&=\; -\lambda\Big[\frac{\pi}{2\gamma}+\frac{r(\eta,\beta)\delta-\pi}{(r(\eta,\beta)+\nu)(T-s)+2\gamma}\Big]ds \\ &\quad + \frac{r(\eta,\beta)\sigma_d}{(r(\eta,\beta)+\nu)(T-s)+2\gamma}dB_s \;-\; \frac{\sigma_0}{(r(\eta,\beta)+\nu)(T-s)+2\gamma}dW_s \\ &\quad + \frac{r(\eta,\beta)\delta^+-\pi^+}{(r(\eta,\beta)+\nu)(T-s)+2\gamma}dN_s^+ + \frac{r(\eta,\beta)\delta^--\pi^-}{(r(\eta,\beta)+\nu)(T-s)+2\gamma}dN_s^-, \end{aligned}$$

where we write the dynamics directly in terms of the Poisson processes $N^{\pm}$. By integration, we deduce the (path-dependent) expression of $\hat{q}_s^{(\lambda)}$, $t \leq s \leq T$:

$$\begin{aligned}
\hat{q}_s^{(\lambda)} &= \hat{q}_t^{(\lambda)} - \frac{\lambda\pi}{2\gamma}(s-t) + \frac{\lambda\big(r(\eta,\beta)\delta - \pi\big)}{r(\eta,\beta)+\nu} \ln\Big(\frac{(r(\eta,\beta)+\nu)(T-s)+2\gamma}{(r(\eta,\beta)+\nu)(T-t)+2\gamma}\Big) \\
&\quad + \int_t^s \frac{r(\eta,\beta)\sigma_d}{(r(\eta,\beta)+\nu)(T-u)+2\gamma} dB_u - \int_t^s \frac{\sigma_0}{(r(\eta,\beta)+\nu)(T-u)+2\gamma} dW_u \\
&\quad + \int_t^s \frac{r(\eta,\beta)\delta^+ - \pi^+}{(r(\eta,\beta)+\nu)(T-u)+2\gamma} dN_u^+ + \int_t^s \frac{r(\eta,\beta)\delta^- - \pi^-}{(r(\eta,\beta)+\nu)(T-u)+2\gamma} dN_u^-,
\end{aligned}$$

with $\hat{q}_t^{(\lambda)} = \hat{q}^{(\lambda)}(T-t, d-x, y)$. We thus obtain the expression of the final spread between demand and inventory:

$$\begin{aligned}
D_T^{t,d} - \hat{X}_T^{t,x,y,d} =& d - x + \mu(T-t) + \int_t^T \sigma_d dB_s + \int_t^T \delta^+ dN_s^+ + \int_t^T \delta^- dN_s^- - \int_t^T \hat{q}_s^{(\lambda)} ds \\
=& m^{(\lambda)}(T-t, d-x, y) \\
&+ \int_t^T \frac{\sigma_d(\nu(T-s)+2\gamma)}{(r(\eta,\beta)+\nu)(T-s)+2\gamma} dB_s + \int_t^T \frac{\sigma_0(T-s)}{(r(\eta,\beta)+\nu)(T-s)+2\gamma} dW_s \\
&+ \int_t^T \frac{\delta^+(\nu(T-s)+2\gamma) + \pi^+(T-s)}{(r(\eta,\beta)+\nu)(T-s)+2\gamma} dN_s^+ \\
&+ \int_t^T \frac{\delta^-(\nu(T-s)+2\gamma) + \pi^-(T-s)}{(r(\eta,\beta)+\nu)(T-s)+2\gamma} dN_s^-,
\end{aligned} \tag{4.46}$$

by Fubini's theorem, and where

$$\begin{aligned}
m^{(\lambda)}(t,d,y) &:= d + \mu t - t\hat{q}^{(\lambda)}(t,d,y) + \frac{\lambda\pi}{2\gamma}\int_0^t s ds \\
&\quad - \frac{\lambda\big(r(\eta,\beta)\delta - \pi\big)}{r(\eta,\beta)+\nu} \int_0^t \ln\Big(\frac{(r(\eta,\beta)+\nu)s+2\gamma}{(r(\eta,\beta)+\nu)t+2\gamma}\Big) ds,
\end{aligned}$$

is explicitly written as in (4.44) after some straightforward calculation. Denoting by $\Delta_T^{t,x,y,d}$ the continuous part of $D_T^{t,d} - \hat{X}_T^{t,x,y,d}$ consisting in the three first terms in the rhs of (4.46), and by $\Sigma_T^{+,t}$, $\Sigma_T^{-,t}$ the jump parts consisting in the two last terms of (4.46), so that

$$D_T^{t,d} - \hat{X}_T^{t,x,y,d} = \Delta_T^{t,x,y,d} + \Sigma_T^{+,t} + \Sigma_T^{-,t},$$

we notice that $\Delta_T^{t,x,y,d}$ follows a normal distribution law with mean $m^{(\lambda)}(T-t, d-x, y)$ and variance $V(T-t)$, independent of $\Sigma_T^{\pm,t}$. Then, conditionally on $\Sigma_T^{\pm,t}$, $D_T^{t,d} - \hat{X}_T^{t,x,y,d}$ follows a normal distribution law with mean $m^{(\lambda)}(T-t, d-x, y) + \Sigma_T^{+,t} + \Sigma_T^{-,t}$, and variance $V(T-t)$, and this implies from (4.45) that:

$$\begin{aligned}
\mathcal{E}^{(\lambda)}(t,x,y,d) &= \frac{\eta r(\eta,\beta)}{2\beta} V(T-t)\mathbb{E}\Big[\psi\Big(\frac{m^{(\lambda)}(T-t,d-x,y) + \Sigma_T^{+,t} + \Sigma_T^{-,t}}{\sqrt{V(T-t)}}\Big)\Big] \\
&\leq \frac{\eta r(\eta,\beta)}{2\beta} V(T-t)\mathbb{E}\Big[\psi\Big(\frac{m^{(\lambda)}(T-t,d-x,y) + \Sigma_T^{-,t}}{\sqrt{V(T-t)}}\Big)\Big],
\end{aligned}$$

since $\Sigma_T^{+,t} \geq 0$ a.s. and ψ is non-increasing. $\square$

Comments on the approximation error. Let us discuss about the accuracy of the upper bound in (4.43) :

$$\bar{\mathcal{E}}^{(\lambda)}(T-t, d-x, y) \quad := \quad \frac{\eta r(\eta,\beta)}{2\beta} V(T-t) \mathbb{E}\Big[\psi\Big(\frac{m^{(\lambda)}(T-t, d-x, y) + \Sigma_T^{-,t}}{\sqrt{V(T-t)}}\Big)\Big],$$

First, notice that $m^{(\lambda)}(T-t, d-x, y) + \Sigma_T^{-,t} \sim m(T-t, d-x, y)$ a.s. in the limiting regimes where $T-t$ goes to zero, $d-x$ or y goes to infinity. Therefore, by dominated convergence theorem, $\bar{\mathcal{E}}^{(\lambda)}(T-t, d-x, y)$ converges to zero in these limiting regimes as in the no jump case. However, we are not able to derive an asymptotic limit as in the no jump case of Proposition 4.3, except when $\Sigma_T^{-,t} = 0$, i.e. $\delta^- = \pi^- = 0$, for which we get the same asymptotic limit. Actually, in the presence of negative jumps on the demand, it is intuitively clear that our approximation should be less accurate than in the no jump case since the probability for the residual demand to stay above the final inventory is decreasing. Anyway, the explicit strategies $(\hat{q}^{(\lambda)}, \tilde{\xi}_T^{(\lambda),*})$ still provide a very accurate approximation of the optimal strategies at least in these limiting regimes, as illustrated in the next paragraph.

4.4.3 Numerical results

We plot trajectories of some relevant quantities that we simulate with the same set of parameters as in Paragraph 4.3.3 : $\sigma_0 = 1/60$ €$\cdot$(MW)$^{-1} \cdot s^{-1/2}$, $\sigma_d = 1000/60$ MW$\cdot s^{-1/2}$, $\beta = 0.002$ €$\cdot$(MW)$^{-2}$, $\eta = 200$ €$\cdot$(MW)$^{-2}$, $\mu = 0$ MW$\cdot s^{-1}$, $\rho = 0.8$, $\nu = 4.00 \cdot 10^{-5}$€$\cdot$(MW)$^{-2}$, $\gamma = 2.22$€$\cdot s \cdot$(MW)$^{-2}$, $T = 24$h, $X_0 = 0$, $D_0 = 50,000$ MW and $Y_0 = 50$ €$\cdot$(MW)$^{-1}$. Moreover, we fix the probability of positive jumps, $p^+ = 1$ (then all jumps are positive: $p^- = 0$), and the following values for the jump components: $\lambda = 1.5/(3600 \cdot 24)\ s^{-1}$, $\pi^+ = 10$ €$\cdot$(MW)$^{-1}$, $\delta^+ = 1500$ MW.
For such parameter values, we observe two occurrences of jumps on the trajectories of the demand of price. Moreover, the probability $\mathbb{P}[\hat{X}_T > D_T]$ is bounded above by 2.92×10^{-16}, the error $\bar{\mathcal{E}}^{(\lambda)}(0, D_0 - X_0, Y_0)$ is bounded by 2.66×10^{-5}€, and

$$\tilde{v}^{(\lambda)}(0, X_0, Y_0, D_0) \quad = \quad 2020950€.$$

The executed strategy $(\hat{q}^{(\lambda)}, \hat{\xi}_T^{(\lambda),*})$ can then be considered as very close to the optimal strategy. This has to be compared with the numerical result obtained in the previous section in the no jump case where we obtained a lower expected total cost: $\tilde{v}(0, X_0, Y_0, D_0) = 1916700$€. Figure 4.4 represents the evolution of the trading rate $(\hat{q}_t^{(\lambda)})_{t\in[0,T]}$, and we see that it is decreasing consistently with the supermartingale property in Proposition 4.4. Actually, we observe that the deterministic part in (4.40), which is linear in time, dominates the stochastic part. The interpretation of the strategy is the following: since positive price jumps are expected, the agent purchases a large number of shares in electricity with the hope to sell it later at a higher price thanks to the possible occurrence of a positive jump. At the price jump times, which can be visualized in Figure 4.5, we notice that the control $\hat{q}^{(\lambda)}$ reacts by a decrease in the trading rate. The reaction to the second jump is more sensible than to the first jump since it occurs a short time before the final horizon T, where the objective is also to achieve the equilibrium relation (4.41) between price and marginal cost. Finally, we observe clearly in Figure 4.6 the concavity of the trajectory of the optimal inventory process $(\hat{X}_t)_{t\in[0,T)}$, as expected from Remark 4.5. This emphasizes the double objective of the agent: on one hand, the purchase of electricity shares for taking profit of the positive price jumps, and on the other hand the resale of electricity shares for attaining the equilibrium relation between price and marginal cost at terminal date. We also plot the production $\hat{\xi}_T$ at the final time T in Figure 4.6, and observe as in the no jump case that the imbalance cost $D_T - \hat{X}_T - \hat{\xi}_T$ is positive.

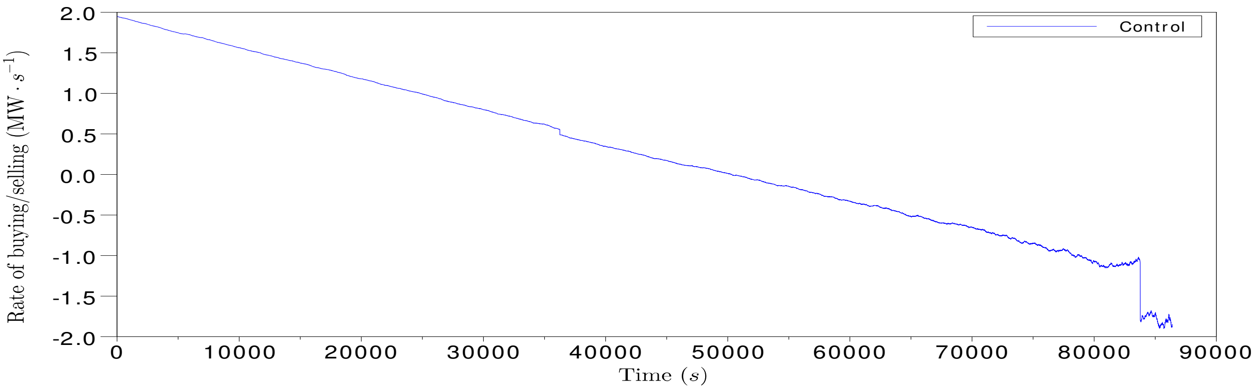

Figure 4.4 – Evolution of the trading rate control $\hat{q}^{(\lambda)}$

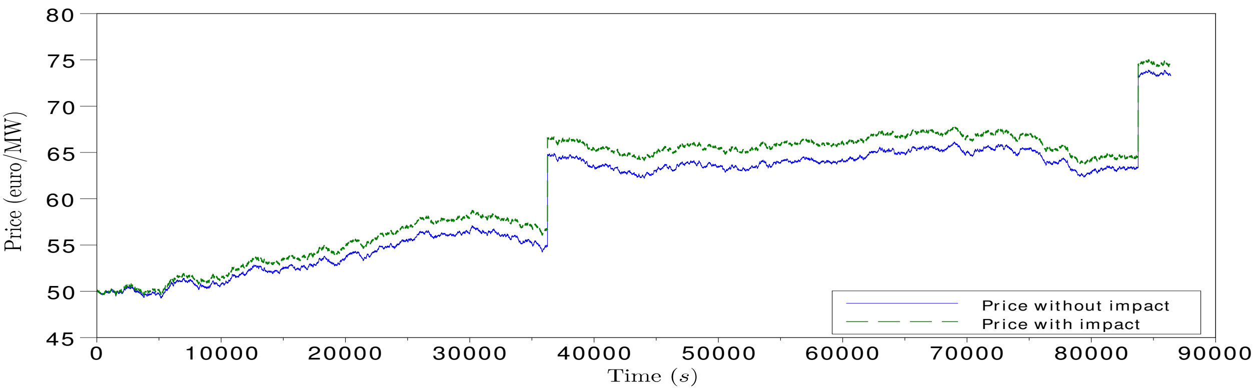

Figure 4.5 – Simulation of the quoted impacted price $\hat{Y}$ and of the unaffected price $\hat{P}$

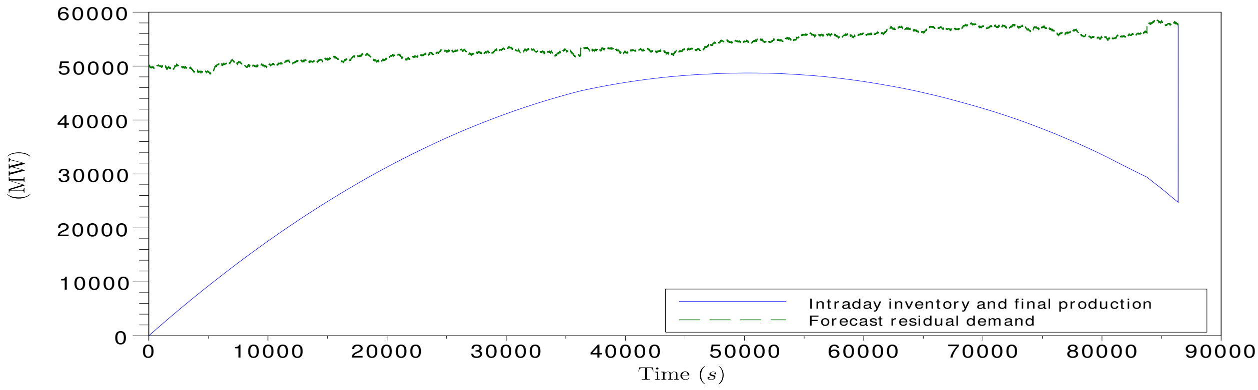

Figure 4.6 – Evolution of the inventory $\hat{X}$ and of the forecast residual demand D

Next, we plot trajectories with the same set of parameters, but with $p^+ = 0.3$ (i.e. $p^- = 0.7$), $\pi^- = -10$ €$\cdot$(MW)$^{-1}$, $\delta^- = -1500$ MW. There are, in average, more negative than positive jumps. Now

$$\tilde{v}^{(\lambda)}(0, X_0, Y_0, D_0) \quad = \quad 1756330\text{€}.$$

Figure 4.7 shows that the trading rate $(\hat{q}_t^{(\lambda)})_{t\in[0,T]}$ is increasing, which is consistent with the submartingale property in Proposition 4.4 : the deterministic part in (4.40) dominates the stochastic part. Since negative jumps are more expected than negative jumps are, the agent first sells a large number of shares in electricity with the hope to buy it later at a lower price thanks to the possible occurrence of jumps, that should be mainly negative. Here, the control reacts to the negative price jumps by an increase in the trading rate. Finally, in Figure 4.9 we observe the convexity of the trajectory of the optimal inventory $(\hat{X}_t)_{t\in[0,T)}$ process, as expected from Remark 4.5. We also plot the production $\hat{\xi}_T$ at the final time T in that figure.

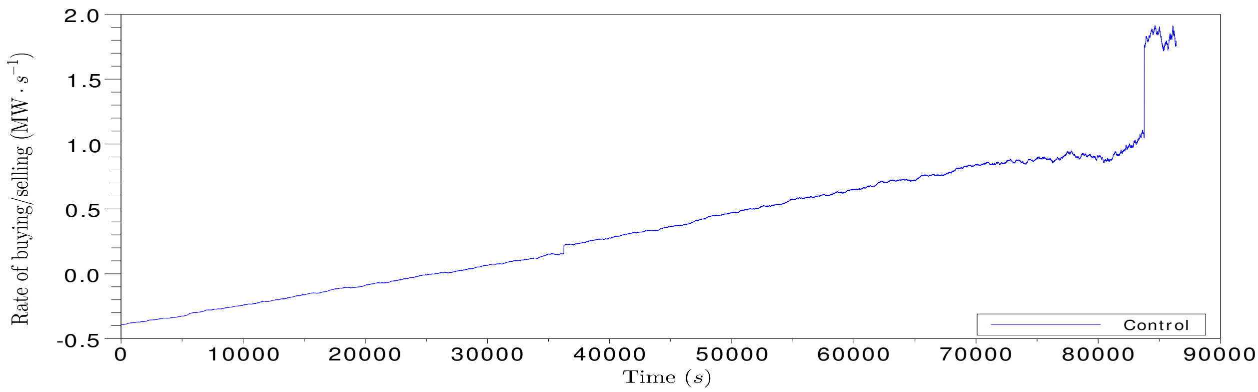

Figure 4.7 – Evolution of the trading rate control $\hat{q}^{(\lambda)}$

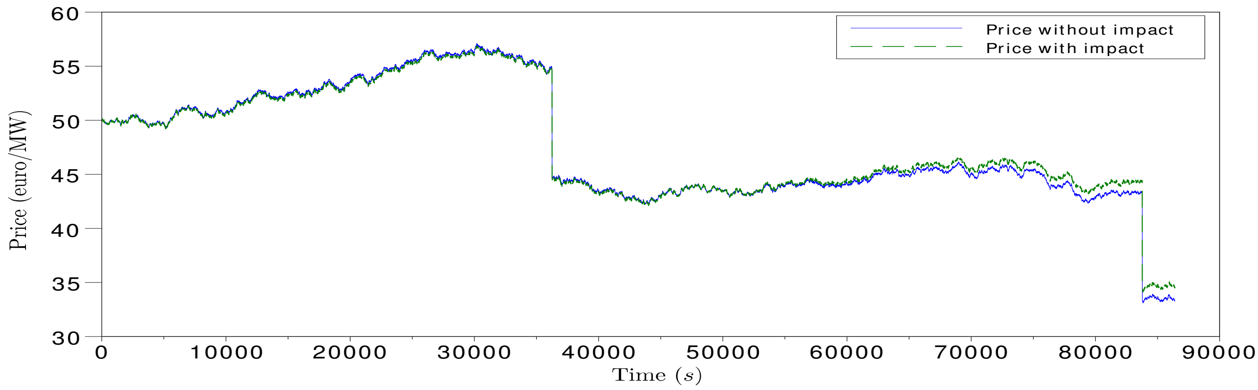

Figure 4.8 – Simulation of the quoted impacted price $\hat{Y}$ and of the unaffected price $\hat{P}$

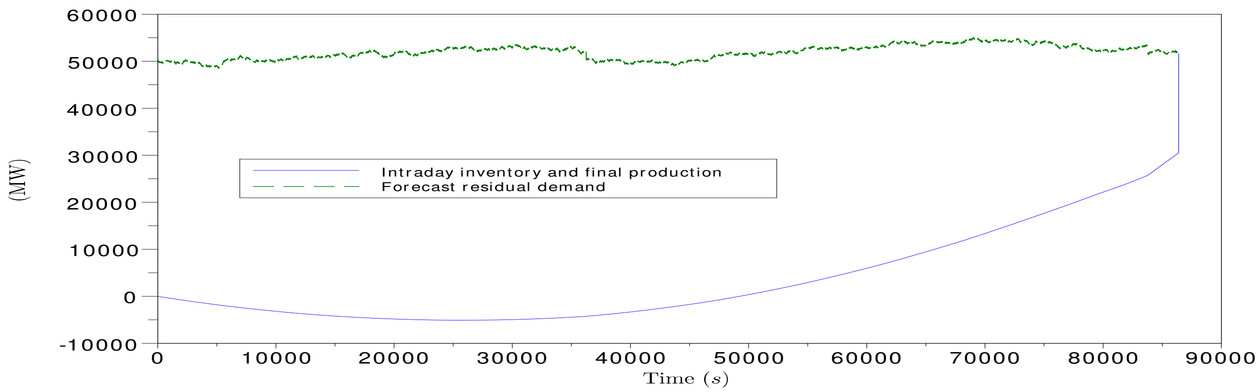

Figure 4.9 – Evolution of the inventory $\hat{X}$ and of the forecast residual demand D

4.5 Delay in production

In this section, we consider the more realistic situation when there is delay in the production, assumed to be fixed equal to $h \in [0, T]$, and we denote by $v = v_h$ the value function to the associated optimal execution problem, as defined in (4.9), where we stress the dependence in the delay h. Our aim is to show how one can reduce the problem with delay to a suitable problem without delay, and then solve it explicitly. We shall consider the problem without jumps on demand forecast and price, but the same argument also works for the case with jumps.

4.5.1 Explicit solution with delay

For simplicity of presentation, and without loss of generality, we shall focus on the derivation of the value function $v_h(t, x, y, d)$ for an initial time $t = 0$, and fixed $(x, y, d) \in \mathbb{R} \times \mathbb{R} \times \mathbb{R}$. Given a control trading rate $q \in \mathcal{A}$, and from pathwise uniqueness for the solution to the dynamics (4.1), (4.7), (4.8), we observe that for any $\xi \in L^0(\mathcal{F}_{T-h})$:

$$\begin{cases} X_T^{0,x} + \xi \ = \ X_T^{T-h, X_{T-h}^{0,x}+\xi} \ a.s. \\ Y_T^{0,y} \ = \ Y_T^{T-h, Y_{T-h}^{0,y}}, \quad D_T^{0,d} \ = \ D_T^{T-h, D_{T-h}^{0,d}} \ a.s. \end{cases} \tag{4.47}$$

To alleviate notations, we shall omit the dependence in the fixed initial conditions (x, y, d), and simply write $X_s = X_s^{0,x}$, $Y_s = Y_s^{0,y}$, $D_s = D_s^{0,d}$, for $s \geq 0$, $v_h = v_h(0, x, y, d)$, and $J(0; q, \xi)$ $= J(0, x, y, d; q, \xi)$ for the the cost functional in (4.10). By the tower property of conditional expectations and from (4.47), the cost functional can be written, for all $q \in \mathcal{A}$, $\xi \in L^0(\mathcal{F}_{T-h})$, as:

$$\begin{aligned} & J(0; q, \xi) \\ = \ & \mathbb{E}\Big[\int_0^{T-h} q_s\big(Y_s + \gamma q_s\big)ds + c(\xi) + J(T-h, X_{T-h} + \xi, Y_{T-h}, D_{T-h}; q, 0)\Big] \quad (4.48) \\ \geq \ & \mathbb{E}\Big[\int_0^{T-h} q_s\big(Y_s + \gamma q_s\big)ds + c(\xi) + v_{_{NP}}(T-h, X_{T-h} + \xi, Y_{T-h}, D_{T-h})\Big], \end{aligned}$$

by definition (4.18) of the value function $v_{_{NP}}$ for the optimal execution problem without production, i.e. the pure retailer problem. Since q is arbitrary in $\mathcal{A}$, this shows that:

$$\begin{aligned} & \inf_{q \in \mathcal{A}} J(0; q, \xi) \quad (4.49) \\ \geq \ & \inf_{q \in \mathcal{A}} \mathbb{E}\Big[\int_0^{T-h} q_s\big(Y_s + \gamma q_s\big)ds + c(\xi) + v_{_{NP}}(T-h, X_{T-h} + \xi, Y_{T-h}, D_{T-h})\Big], \end{aligned}$$

for all $\xi \in L^0(\mathcal{F}_{T-h})$. Now, given $q \in \mathcal{A}$, and $\xi \in L^0(\mathcal{F}_{T-h})$, let us consider the trading rate $\hat{q}^{NP,\xi}$ in $\mathcal{A}_{T-h}$ solution to the pure retailer problem: $v_{_{NP}}(T-h, X_{T-h}+\xi, Y_{T-h}, D_{T-h})$, hence starting at time $T-h$ from an inventory $X_{T-h}+\xi$. By considering the process $\tilde{q} \in \mathcal{A}$ defined by: $\tilde{q}_s = q_s$ for $0 \leq s < T-h$, and $\tilde{q}_s = \hat{q}_s^{NP,\xi}$, for $T-h \leq s \leq T$, we then obtain from (4.48) :

$$\begin{aligned} & J(0;\tilde{q},\xi) \\ = \; & \mathbb{E}\Big[\int_0^{T-h} q_s\big(Y_s+\gamma q_s\big)ds + c(\xi) + v_{NP}(T-h, X_{T-h}+\xi, Y_{T-h}, D_{T-h})\Big], \end{aligned} \quad (4.50)$$

which proves together with (4.49) the equality:

$$\begin{aligned} & \inf_{q\in\mathcal{A}} J(0;q,\xi) \\ = \; & \inf_{q\in\mathcal{A}} \mathbb{E}\Big[\int_0^{T-h} q_s\big(Y_s+\gamma q_s\big)ds + c(\xi) + v_{NP}(T-h, X_{T-h}+\xi, Y_{T-h}, D_{T-h})\Big], \end{aligned} \quad (4.51)$$

for all $\xi \in L^0(\mathcal{F}_{T-h})$. Therefore, $v_h = \inf_{q\in\mathcal{A},\xi\in L^0_+(\mathcal{F}_{T-h})} J(0;q,\xi)$ can be written as:

$$\begin{aligned} v_h \;=\; & \inf_{q\in\mathcal{A},\xi\in L^0_+(\mathcal{F}_{T-h})} \mathbb{E}\Big[\int_0^{T-h} q_s\big(Y_s+\gamma q_s\big)ds \\ & \qquad + c(\xi) + v_{NP}(T-h, X_{T-h}+\xi, Y_{T-h}, D_{T-h})\Big]. \end{aligned} \quad (4.52)$$

In other words, the original problem with delay in production is formulated as an optimal execution problem without delay, namely with final horizon $T-h$, and terminal cost function:

$$C_h(x,y,d,\xi) \;:=\; c(\xi) + v_{NP}(T-h, x+\xi, y, d).$$

Notice from the explicit expression of v_{NP} in Remark 4.3 that this cost function C_h does not depend on T, and is in the form:

$$C_h(x,y,d,\xi) \;=\; C_h(0,y,d-x-\xi,0) \;=\; c(\xi) + v_{NP}(T-h,0,y,d-x-\xi).$$

The optimization over q and ξ in (4.52) is done separately: the production $\xi \in L^0_+(\mathcal{F}_{T-h})$ is decided at time $T-h$, after the choice of the trading rate (q_s) for $0 \leq s \leq T-h$ (leading to an inventory X_{T-h}), and is determined optimally from the optimization a.s. at $T-h$ of the terminal cost $C_h(X_{T-h}, Y_{T-h}, D_{T-h}, \xi)$. It is then given in feedback form by $\xi^*_{T-h} = \hat{\xi}^{h,tr+}(D_{T-h}-X_{T-h}, Y_{T-h})$ where

$$\hat{\xi}^{h,tr+}(d,y) \;:=\; \arg\min_{\xi\geq 0} C_h(0,y,d-\xi,0) \;=\; \arg\min_{\xi\geq 0} \big[c(\xi) + v_{NP}(T-h,0,y,d-\xi)\big],$$

hence explicitly given from the expression of v_{NP} in Remark 4.3 by:

$$\begin{aligned} \hat{\xi}^{h,tr+}(d,y) \;&=\; \hat{\xi}^h(d,y)\mathbf{1}_{\hat{\xi}^h(d,y)\geq 0}, \\ \hat{\xi}^h(d,y) \;&:=\; \frac{\eta}{\eta+\beta}\Big[\frac{(\nu h+2\gamma)(\mu h+d)+hy}{(r(\eta,\beta)+\nu)h+2\gamma}\Big]. \end{aligned} \quad (4.53)$$

The problem (4.52) is then rewritten as

$$v_h = \inf_{q\in\mathcal{A}} \mathbb{E}\Big[\int_0^{T-h} q_s\big(Y_s+\gamma q_s\big)ds + C_h^+(D_{T-h}-X_{T-h},Y_{T-h})\Big], \tag{4.54}$$

where

$$C_h^+(d,y) \;:=\; C_h(0,y,d-\hat{\xi}^{h,tr+}(d),0).$$

Notice that when $h=0$, we retrieve the expressions in the no delay case: $\hat{\xi}^{0,tr+}=\hat{\xi}^{tr+}$ in (4.11), $C_0^+=C^+$ in (4.13) and $v_0=v$ in (4.12). As in the no delay case, there is no explicit solution to the HJB equation associated to the stochastic control problem (4.54). We then consider the approximate control problem where we relax the non-negativity constraint on the production, i.e. $\tilde{v}_h=\inf_{q\in\mathcal{A},\xi\in L^0(\mathcal{F}_{T-h})} J(0;q,\xi)$. Therefore by following the same arguments as above, the corresponding value function is written as:

$$\tilde{v}_h = \inf_{q\in\mathcal{A}} \mathbb{E}\Big[\int_0^{T-h} q_s\big(Y_s+\gamma q_s\big)ds + \tilde{C}_h(D_{T-h}-X_{T-h},Y_{T-h})\Big], \tag{4.55}$$

where

$$\tilde{C}_h(d,y) \;:=\; C_h(0,y,d-\hat{\xi}^{h}(d),0).$$

From the explicit expressions of $\hat{\xi}^h$ in (4.53) and v_{NP} in Remark 4.3, it appears after some tedious but straightforward calculations that the auxiliary terminal cost function $\tilde{C}_h$ simplifies remarkably into:

$$\tilde{C}_h(d,y) \;=\; \tilde{v}_0(T-h,0,y,d)+K_h,$$

where $\tilde{v}_0$ is the auxiliary value function without delay explicitly obtained in Theorem 4.1, and K_h is a constant depending only on the delay h and the parameters of the model, given explicitly by

$$\begin{aligned} K_h \;=\; & \frac{\eta^2}{2}\frac{\sigma_0^2+\sigma_d^2\nu^2+2\rho\sigma_0\sigma_d\nu}{(\eta+\beta)(\eta+\nu)(r(\eta,\beta)+\nu)}h \\ & +\gamma\frac{\sigma_0^2+\sigma_d^2\eta^2-2\rho\sigma_0\sigma_d\eta}{(\eta+\nu)^2}\ln\Big(1+\frac{(\eta+\nu)h}{2\gamma}\Big) \\ & -\gamma\frac{\sigma_0^2+\sigma_d^2r^2(\eta,\beta)-2\rho\sigma_0\sigma_d r(\eta,\beta)}{(r(\eta,\beta)+\nu)^2}\ln\Big(1+\frac{(r(\eta,\beta)+\nu)h}{2\gamma}\Big). \end{aligned}$$

One easily checks that $K_h=0$ for $h=0$, and K_h is increasing with h (actually the derivative of K_h w.r.t. h is positive), hence in particular K_h is nonnegative. Plugging into (4.55), we then get

$$\tilde{v}_h = \inf_{q\in\mathcal{A}} \mathbb{E}\Big[\int_0^{T-h} q_s\big(Y_s+\gamma q_s\big)ds + \tilde{v}_0(T-h,X_{T-h},Y_{T-h},D_{T-h})\Big] + K_h. \tag{4.56}$$

Therefore, by using the dynamic programming principle for the control problem

$$\tilde{v}_0 = \tilde{v}_0(0, x, y, d)$$

in (4.14), we obtain this remarkable relation

$$\tilde{v}_h \;=\; \tilde{v}_0 \;+\; K_h, \tag{4.57}$$

which explicitly relates the (approximate) value function with and without delay. As expected from the very definition of $\tilde{v}_h$, this relation implies that $\tilde{v}_h - \tilde{v}_0$ is nonnegative, and is increasing in h. This is consistent with the intuition that when making the production choice in advance, we do not take into account the future movements of the price and of the residual demand, which should therefore lead to an average positive correction of the cost. More precisely, the relation (4.57) gives an explicit quantification of the delay impact via the term K_h (which does not depend on the state variables x, y, d) in function of the various model parameters. Moreover, the optimal control of the stochastic control problem (4.56) over $[0, T-h)$ is explicitly given by the optimal control $(\hat{q}_s)_{0\leq s\leq T-h}$ of problem $\tilde{v}_0$ without delay in Theorem 4.1.

Let us now consider the following strategy $(\hat{q}^{h,+}, \tilde{\xi}^{h,*}_{T-h}) \in \mathcal{A} \times L^0_+(\mathcal{F}_{T-h})$ for the original problem v_h with delay:

— Before $T-h$, follow the trading strategy $\hat{q}^{h,+}_s = \hat{q}_s$, $s < T-h$, corresponding to the solution of the auxiliary problem without delay as if production choice is made at time T, and leading to an inventory $\hat{X}_{T-h}$, and an impacted price $\hat{Y}_{T-h}$.

— At time $T-h$, choose the production quantity:

$$\tilde{\xi}^{h,*}_{T-h} \;:=\; \hat{\xi}^{h,tr+}(D_{T-h} - \hat{X}_{T-h}, \hat{Y}_{T-h}).$$

— Between time $T-h$ and T, follow the trading strategy $\hat{q}^{h,+}_s = \hat{q}_s^{NP,\tilde{\xi}^{h,*}_{T-h}}$, $T-h \leq s \leq T$, corresponding to the solution of the problem without production, and starting at $T-h$ from an inventory $\hat{X}_{T-h} + \tilde{\xi}^{h,*}_{T-h}$.

In order to estimate the quality of this approximate strategy with respect to the optimal trading problem v_h, measured by

$$\mathcal{E}^h_1 \;:=\; J(0; \hat{q}^{h,+}, \tilde{\xi}^{h,*}_{T-h}) - v_h,$$

we shall compare it with the following strategy $(\hat{q}^h, \hat{\xi}^h_{T-h}) \in \mathcal{A} \times L^0(\mathcal{F}_{T-h})$:

— Before $T-h$, follow the trading strategy $\hat{q}^h_s = \hat{q}_s$, $s < T-h$, corresponding to the solution of the auxiliary problem without delay as if production choice is made at time T, and leading to an inventory $\hat{X}_{T-h}$, and an impacted price $\hat{Y}_{T-h}$.

— At time $T-h$, choose the "production" quantity (which can be negative):

$$\hat{\xi}^h_{T-h} \;=\; \hat{\xi}^h(D_{T-h} - \hat{X}_{T-h}, \hat{Y}_{T-h}).$$

— Between time $T-h$ and T, follow the trading strategy $\hat{q}^h_s = \hat{q}_s^{NP,\hat{\xi}^h_{T-h}}$, $T-h \leq s \leq T$, corresponding to the solution of the problem without production, and starting at $T-h$ from an inventory $\hat{X}_{T-h} + \hat{\xi}^h_{T-h}$.

Then, by construction and following the arguments (see in particular (4.50), (4.55), (4.56)) leading to the expression (4.57) of $\tilde{v}_h$, we see that $(\hat{q}^h, \hat{\xi}^h_{T-h})$ is the optimal solution for $\tilde{v}_h$, i.e. $\tilde{v}_h = J(0; \hat{q}^h, \hat{\xi}^h_{T-h})$. On the other hand, since $\tilde{v}_h \leq v_h \leq J(0; \hat{q}^{h,+}, \tilde{\xi}^{h,*}_{T-h})$, we deduce that

$$\max(v_h - \tilde{v}_h, \mathcal{E}^h_1) \quad \leq \quad \bar{\mathcal{E}}^h \; := \; J(0; \hat{q}^{h,+}, \tilde{\xi}^{h,*}_{T-h}) - J(0; \hat{q}^h, \hat{\xi}^h_{T-h}).$$

Now, from the expression (4.50) of J, and by same arguments as in the proof of Proposition 4.2 (see the derivation of relation (4.25)), we have

$$\begin{aligned} \bar{\mathcal{E}}^h \;&=\; \mathbb{E}\Big[C^+_h(D_{T-h} - \hat{X}_{T-h}, \hat{Y}_{T-h}) - \tilde{C}_h(D_{T-h} - \hat{X}_{T-h}, \hat{Y}_{T-h})\Big] \\ &=\; \mathbb{E}\Big[v_{NP}(T-h, 0, \hat{Y}_{T-h}, D_{T-h} - \hat{X}_{T-h} - \tilde{\xi}^{h,*}_{T-h}) + c(\tilde{\xi}^{h,*}_{T-h}) \\ &\qquad - v_{NP}(T-h, 0, \hat{Y}_{T-h}, D_{T-h} - \hat{X}_{T-h} - \hat{\xi}^h_{T-h}) - c(\hat{\xi}^h_{T-h})\Big] \\ &=\; \frac{\eta r(\eta,\beta)}{2\beta} \frac{(r(\eta,\beta)+\nu)h + 2\gamma}{(\eta+\nu)h + 2\gamma} V_h(T) \psi\Big(\frac{m(T, d-x, y)}{\sqrt{V_h(T)}}\Big) \end{aligned}$$

where m and ψ are defined as in (4.22), and

$$V_h(T) \;=\; \int_h^T \frac{\sigma_0^2 s^2 + \sigma_d^2(\nu s + 2\gamma)^2 + 2\rho\sigma_0\sigma_d s(\nu s + 2\gamma)}{\big[(r(\eta,\beta)+\nu)s + 2\gamma\big]^2} ds.$$

We recover when $h = 0$ the expression in Proposition 4.2 of the error in the no delay case, and notice that $\bar{\mathcal{E}}^h$ decreases when the delay increases: indeed, the error comes from the trading procedure before deciding how much to produce, which is dictated by the auxiliary problem, in which the final “production” can be negative. After $T-h$, the followed control is optimal, as there remains no production decision at some further date. The shorter the period before making the production decision is, the weaker the error is.

Let us finally discuss some properties of the (approximate) optimal trading strategy $\hat{q}^{h,+}$. Recalling from Proposition 4.1 that the optimal trading rate is a martingale in the no delay case, we see by construction of $(\hat{q}^{h,+}_s)_{0 \leq s \leq T}$ that it is a martingale on $[0, T-h)$ and a martingale on $[T-h,T]$. Moreover, for any $s \in [T-h,T]$, and $t \in [0,T-h)$, we have

$$
\begin{aligned}
\mathbb{E}\big[\hat{q}_s^{h,+}|\mathcal{F}_t\big] &= \mathbb{E}\big[\mathbb{E}\big[\hat{q}_s^{NP,\tilde{\xi}^{h,*}_{T-h}}|\mathcal{F}_{T-h}\big]|\mathcal{F}_t\big] \;=\; \mathbb{E}\big[\hat{q}_{T-h}^{NP,\tilde{\xi}^{h,*}_{T-h}}|\mathcal{F}_t\big] \\
&= \mathbb{E}\Big[\frac{\eta(\mu h + D_{T-h} - \hat{X}_{T-h} - \tilde{\xi}^{h,*}_{T-h}) - \hat{Y}_{T-h}}{(\eta+\nu)h + 2\gamma}\big|\mathcal{F}_t\Big] \\
&= \mathbb{E}\Big[\frac{\eta(\mu h + D_{T-h} - \hat{X}_{T-h} - \hat{\xi}^{h}_{T-h}) - \hat{Y}_{T-h}}{(\eta+\nu)h + 2\gamma}\big|\mathcal{F}_t\Big] \\
&\quad + \frac{\eta}{(\eta+\nu)h+2\gamma}\mathbb{E}\Big[\hat{\xi}^h_{T-h} - \tilde{\xi}^{h,*}_{T-h}\big|\mathcal{F}_t\Big] \\
&= \mathbb{E}\Big[\frac{r(\eta,\beta)(\mu h + D_{T-h} - \hat{X}_{T-h}) - \hat{Y}_{T-h}}{(r(\eta,\beta)+\nu)h + 2\gamma}\big|\mathcal{F}_t\Big] \\
&\quad + \frac{\eta}{(\eta+\nu)h+2\gamma}\mathbb{E}\Big[\hat{\xi}^h_{T-h} - \tilde{\xi}^{h,*}_{T-h}\big|\mathcal{F}_t\Big] \\
&= \mathbb{E}\big[\hat{q}_{T-h}|\mathcal{F}_t\big] + \frac{\eta}{(\eta+\nu)h+2\gamma}\mathbb{E}\Big[\hat{\xi}^h_{T-h} - \tilde{\xi}^{h,*}_{T-h}\big|\mathcal{F}_t\Big] \\
&= \hat{q}_t + \frac{\eta}{(\eta+\nu)h+2\gamma}\mathbb{E}\Big[\hat{\xi}^h_{T-h}\mathbf{1}_{\hat{\xi}^h_{T-h}<0}\big|\mathcal{F}_t\Big] \;\le\; \hat{q}_t \;=\; \hat{q}_t^{h,+}.
\end{aligned}
\tag{4.58}
$$

where we used the tower rule for conditional expectations, the martingale property and the explicit expression of $q^{NP,\tilde{\xi}^{h,*}_{T-h}}$ in Remark 4.3, the definition of $\hat{\xi}^h_{T-h}$, the martingale property and explicit expression of $\hat{q}$ in Theorem 4.1, and finally the fact that $\tilde{\xi}^{h,*}_{T-h} = \hat{\xi}^h_{T-h}\mathbf{1}_{\hat{\xi}^h_{T-h}\ge 0}$. This shows in particular the supermartingale property of $\hat{q}^{h,+}$ over the whole period $[0,T]$. Notice that the same arguments as for the derivation of (4.58) shows the martingale property over the whole period $[0,T]$ of the optimal trading strategy $\hat{q}^h$ associated to the auxiliary problem $\tilde{v}_h$. Moreover, by the martingale property of $\hat{q}^{h,+}$ on $[0,T-h)$, and relation (4.58), we see that the (approximate) optimal inventory process $\hat{X}^{h,+}$ with trading rate $\hat{q}^{h,+}$ has on average, a growth rate $\frac{d\mathbb{E}[\hat{X}_s^{h,+}]}{ds}$, which is piecewise constant, equal to:

$$
\mathbb{E}[\hat{q}_s^{h,+}] \;=\; \begin{cases} \hat{q}_0, & \text{for } \; 0 \le s < T-h \\ \hat{q}_0^{(h)} := \hat{q}_0 + \frac{\eta}{(\eta+\nu)h+2\gamma}\mathbb{E}\big[\hat{\xi}^h_{T-h}\mathbf{1}_{\hat{\xi}^h_{T-h}<0}\big] \;<\; \hat{q}_0, & \text{for } \; T-h \le s \le T, \end{cases}
$$

with $\hat{q}_0 = \frac{r(\eta,\beta))(\mu T + d - x) - y}{(r(\eta,\beta)+\nu)T + 2\gamma}$, and

$$
\hat{q}_0^{(h)} \;=\; \hat{q}_0 - \frac{\eta r(\eta,\beta)}{\beta((\eta+\nu)h+2\gamma)}\sqrt{V_h(T)}\tilde{\psi}\Big(\frac{m(T,d-x,y)}{\sqrt{V_h(T)}}\Big),
$$

where

$$
\tilde{\psi}(z) \;:=\; \phi(z) - z\Phi(-z), \quad z \in \mathbb{R}
$$

is a nonnegative function, as pointed out in (4.27).

4.5.2 Numerical results

We plot figures showing relevant trajectories with the same parameters as in Section 4.3.3. We add a delay $h = 4$ hours: the production choice has to be made four hours before the end of the trading period. We have

$$\tilde{v}_h(0, X_0, Y_0, D_0) \quad = \quad 1925460€,$$

which is slightly higher than the value $\tilde{v}_0(0, X_0, Y_0, D_0) = 1916700€$ without delay.

In Figure 4.10, we see that at time $T - h$, the positive production choice $\tilde{\xi}^{h,*}_{T-h}$ is made, and then we go on buying shares on the intraday market in order to go nearer to the demand forecast, with a smaller slope of trading rate. In Figure 4.11, which represents the control process without the last hour of trading (because oscillations then become overwhelming), we see that after date $T - h$, as we do not plan to use final production leverage any more, the approximate optimal control process $\hat{q}^{h,+}$ oscillates a lot as we are approaching the end of trading time. We can compare with Figure 4.1 to assert that qualitatively, the control in the problem with no production oscillates more than the one in the problem with final production, as in the former problem, the intraday market is the only way to seek to reach the equilibrium.

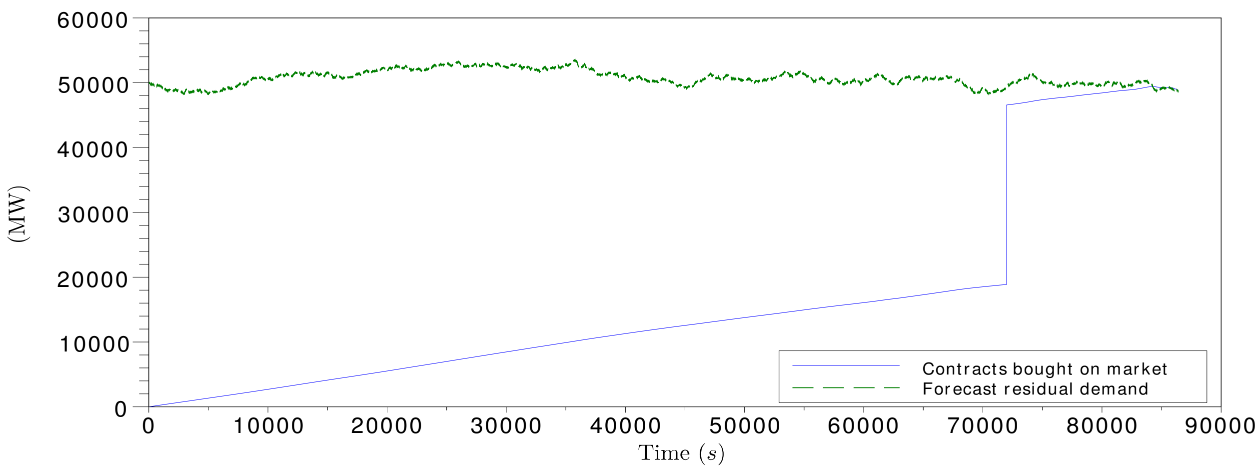

Figure 4.10 – Evolution of the inventory $\hat{X}$ (with production choice at time $T - h$) and of the forecast residual demand D

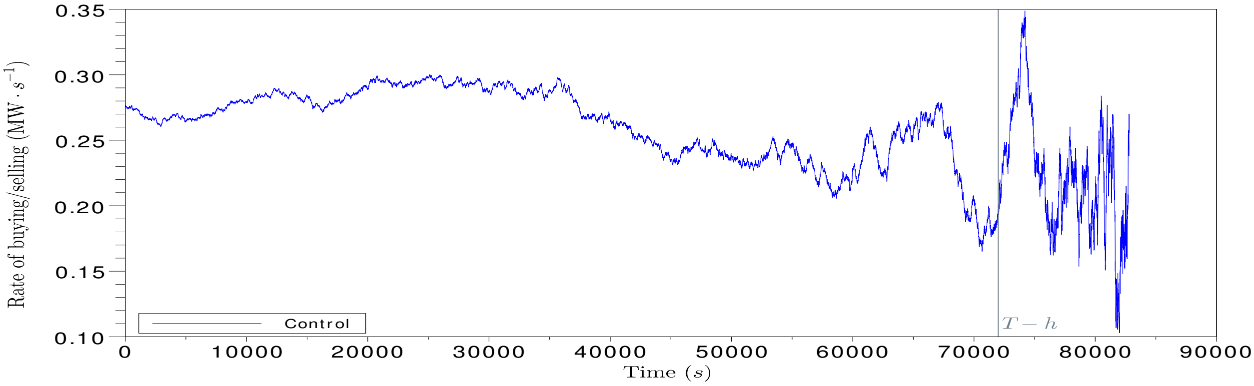

Figure 4.11 – Evolution of the trading rate control $\hat{q}^{h,+}$ without the last hour

4.6 Appendices

4.6.1 Proof of Theorem 4.1

The Hamilton-Jacobi-Bellman (HJB) equation arising from the dynamic programming associated to the stochastic control problem (4.14) is:

$$\begin{cases} \frac{\partial \tilde{v}}{\partial t} + \inf_{q \in \mathbb{R}} \big[q\frac{\partial \tilde{v}}{\partial x} + \nu q \frac{\partial \tilde{v}}{\partial y} + \mu \frac{\partial \tilde{v}}{\partial d} + \frac{1}{2}\sigma_0^2 \frac{\partial^2 \tilde{v}}{\partial y^2} + \frac{1}{2}\sigma_d^2 \frac{\partial^2 \tilde{v}}{\partial d^2} + \rho\sigma_0\sigma_d \frac{\partial^2 \tilde{v}}{\partial y \partial d} + q(y+\gamma q)\big] \;=\; 0, \\ \tilde{v}(T,x,y,d) \;=\; \tilde{C}(d-x) \;=\; \frac{1}{2} r(\eta,\beta)(d-x)^2. \end{cases}$$

The argmin in HJB is attained for

$$\tilde{q}(t,x,y,d) \;=\; -\frac{1}{2\gamma}\big[\frac{\partial \tilde{v}}{\partial x} + \nu \frac{\partial \tilde{v}}{\partial y} + y\big],$$

and the HJB equation is rewritten as:

$$\begin{cases} \frac{\partial \tilde{v}}{\partial t} + \mu \frac{\partial \tilde{v}}{\partial d} + \frac{1}{2}\sigma_0^2 \frac{\partial^2 \tilde{v}}{\partial y^2} + \frac{1}{2}\sigma_d^2 \frac{\partial^2 \tilde{v}}{\partial d^2} + \rho\sigma_0\sigma_d \frac{\partial^2 \tilde{v}}{\partial y \partial d} - \frac{1}{4\gamma}\big[\frac{\partial \tilde{v}}{\partial x} + \nu \frac{\partial \tilde{v}}{\partial y} + y\big]^2 \;=\; 0, \\ \tilde{v}(T,x,y,d) \;=\; \frac{1}{2} r(\eta,\beta)(d-x)^2. \end{cases} \tag{4.59}$$

We look for a candidate solution to HJB in the form

$$\begin{aligned} \tilde{w}(t,x,y,d) \;=\; & A(T-t)(d-x)^2 + B(T-t)y^2 + F(T-t)(d-x)y \\ & + G(T-t)(d-x) + H(T-t)y + K(T-t), \end{aligned} \tag{4.60}$$

for some deterministic functions A, B, F, G, H and K. Plugging the candidate function $\tilde{w}$ into equation (4.59), we see that $\tilde{w}$ is solution to the HJB equation iff the following system of ordinary differential equations (ODEs) is satisfied by A, B, F, G, H and K :

$$\left\{\begin{array}{rcl} A' + \frac{1}{4\gamma}(-2A+\nu F)^2 &=& 0 \\ B' + \frac{1}{4\gamma}(2\nu B - F + 1)^2 &=& 0 \\ F' + \frac{1}{2\gamma}(-2A+\nu F)(2\nu B - F + 1) &=& 0 \\ G' - 2\mu A + \frac{1}{2\gamma}(-2A+\nu F)(-G+\nu H) &=& 0 \\ H' - \mu F + \frac{1}{2\gamma}(2\nu B - F + 1)(-G+\nu H) &=& 0 \\ K' - \mu G - (\sigma_0^2 B + \sigma_d^2 A + \rho\sigma_0\sigma_d F) + \frac{1}{4\gamma}(-G+\nu H)^2 &=& 0 \end{array}\right.$$

with the initial conditions $A(0) = \frac{1}{2}r(\eta,\beta)$, $B(0) = 0$, $F(0) = 0$, $G(0) = 0$, $H(0) = 0$, $K(0) = 0$. We first solve the Riccati system relative to the triple (A, B, F), and obtain:

$$\left\{\begin{array}{lcl} A(t) &=& \frac{r(\eta,\beta)(\frac{\nu}{2}t+\gamma)}{(r(\eta,\beta)+\nu)t+2\gamma}, \\ B(t) &=& -\frac{1}{2}\frac{t}{(r(\eta,\beta)+\nu)t+2\gamma}, \quad F(t) \;=\; \frac{r(\eta,\beta)t}{(r(\eta,\beta)+\nu)t+2\gamma}. \end{array}\right. \tag{4.61}$$

Then we solve the first-order linear system of ODE relative to the pair (G, H), which leads to the explicit solution:

$$G(t) \;=\; 2\mu t A(t), \qquad \text{and} \qquad H(t) \;=\; -2r(\eta,\beta)\mu t B(t). \tag{4.62}$$

Finally, we explicitly obtain K from the last equation:

$$\begin{aligned} K(t) \;&=\; \gamma\frac{\sigma_0^2 + \sigma_d^2 r^2(\eta,\beta) - 2\rho\sigma_0\sigma_d r(\eta,\beta)}{\big(r(\eta,\beta)+\nu\big)^2}\ln\Big(1 + \frac{(r(\eta,\beta)+\nu)t}{2\gamma}\Big) \\ &\quad + \frac{\sigma_d^2 r(\eta,\beta)\nu + 2\rho\sigma_0\sigma_d r(\eta,\beta) - \sigma_0^2}{2\big(r(\eta,\beta)+\nu\big)}t \;+\; \frac{r(\eta,\beta)\mu^2 t^2(\frac{\nu}{2}t+\gamma)}{(r(\eta,\beta)+\nu)t+2\gamma}. \end{aligned} \tag{4.63}$$

By construction, $\tilde{w}$ in (4.60) with A, B, F, G, H and K explicitly given by (4.61)-(4.62)-(4.63), is a smooth solution with quadratic growth condition to the HJB equation (4.59). Moreover, the argmin in HJB equation for $\tilde{w}$ is attained for

$$\begin{aligned} \tilde{q}(t,x,y,d) \;&=\; -\frac{1}{2\gamma}\Big[\frac{\partial\tilde{w}}{\partial x} + \nu\frac{\partial\tilde{w}}{\partial y} + y\Big] \\ &=\; \frac{r(\eta,\beta)(\mu(T-t)+d-x) - y}{(r(\eta,\beta)+\nu)(T-t)+2\gamma} \;=:\; \hat{q}(T-t, d-x, y). \end{aligned}$$

Notice that $\hat{q}$ is linear, and Lipschitz in x, y, d, uniformly in time t, and so given an initial state (x,y,d) at time t, there exists a unique solution $(\hat{X}^{t,x,y,d}, \hat{Y}^{t,x,y,d}, D^{t,d})_{t\leq s\leq T}$ to (4.1)-(4.7)-(4.8) with the feedback control $\hat{q}_s = \hat{q}(T-s, D_s^{t,d} - \hat{X}_s^{t,x,y,d}, \hat{Y}_s^{t,x,y,d})$, which satisfies: $\mathbb{E}[\sup_{t\leq s\leq T}|\hat{X}_s^{t,x,y,d}|^2 + |\hat{Y}_s^{t,x,y,d}|^2 + |D_s^{t,d}|^2] < \infty$. This implies in particular that $\mathbb{E}[\int_t^T|\hat{q}_s|^2 ds] < \infty$, hence $\hat{q} \in \mathcal{A}_t$. We now call on a classical verification theorem (see e.g. Theorem 3.5.2 in [65]), which shows that $\tilde{w}$ is indeed equal to the value function $\tilde{v}$, and $\hat{q}$ is an optimal control. Finally, once the optimal trading rate $\hat{q}$ is determined, the optimal production is obtained from the optimization over $\xi \in \mathbb{R}$ of the terminal cost $C(D_T^{t,d} - \hat{X}_T^{t,x,y,d}, \xi)$, hence given by: $\hat{\xi}_T = \frac{\eta}{\eta+\beta}(D_T^{t,d} - \hat{X}_T^{t,x,y,d})$. □

4.6.2 Proof of Theorem 4.2

The Hamilton-Jacobi-Bellman (HJB) integro-differential equation arising from the dynamic programming associated to the stochastic control problem $\tilde{v} = \tilde{v}^{(\lambda)}$ with jumps in the dynamics of Y and D is:

$$\begin{cases} \dfrac{\partial \tilde{v}^{(\lambda)}}{\partial t} + \inf_{q\in\mathbb{R}} \big[q\dfrac{\partial \tilde{v}^{(\lambda)}}{\partial x} + \nu q \dfrac{\partial \tilde{v}^{(\lambda)}}{\partial y} + \mu \dfrac{\partial \tilde{v}^{(\lambda)}}{\partial d} \\ \qquad + \frac{1}{2}\sigma_0^2 \dfrac{\partial^2 \tilde{v}^{(\lambda)}}{\partial y^2} + \dfrac{1}{2}\sigma_d^2 \dfrac{\partial^2 \tilde{v}^{(\lambda)}}{\partial d^2} + \rho\sigma_0\sigma_d \dfrac{\partial^2 \tilde{v}^{(\lambda)}}{\partial y \partial d} + q(y+\gamma q)\big] \\ + \lambda\big[p^+ \tilde{v}^{(\lambda)}(t,x,y+\pi^+, d+\delta^+) + p^- \tilde{v}^{(\lambda)}(t,x,y+\pi^-,d+\delta^-) - \tilde{v}^{(\lambda)}(t,x,y,d)\big] \;=\; 0 \\ \tilde{v}^{(\lambda)}(T,x,y,d) \;=\; \tilde{C}(d-x) \;=\; \frac{1}{2} r(\eta,\beta)(d-x)^2. \end{cases}$$

Notice that with respect to the no jump case, there is in addition a linear integro-differential term in the HJB equation (which does not depend on the control), and the argmin is attained as in the no jump case for

$$\tilde{q}^{(\lambda)}(t,x,y,d) \;=\; -\frac{1}{2\gamma}\big[\frac{\partial \tilde{v}^{(\lambda)}}{\partial x} + \nu \frac{\partial \tilde{v}^{(\lambda)}}{\partial y} + y\big].$$

The HJB equation is then rewritten as

$$\begin{cases} \dfrac{\partial \tilde{v}^{(\lambda)}}{\partial t} + \mu \dfrac{\partial \tilde{v}^{(\lambda)}}{\partial d} + \dfrac{1}{2}\sigma_0^2 \dfrac{\partial^2 \tilde{v}^{(\lambda)}}{\partial y^2} + \dfrac{1}{2}\sigma_d^2 \dfrac{\partial^2 \tilde{v}^{(\lambda)}}{\partial d^2} + \rho\sigma_0\sigma_d \dfrac{\partial^2 \tilde{v}^{(\lambda)}}{\partial y\partial d} - \dfrac{1}{4\gamma}\big[\dfrac{\partial \tilde{v}^{(\lambda)}}{\partial x} + \nu \dfrac{\partial \tilde{v}^{(\lambda)}}{\partial y} + y\big]^2 \\ + \lambda\big[p^+ \tilde{v}^{(\lambda)}(t,x,y+\pi^+, d+\delta^+) + p^- \tilde{v}^{(\lambda)}(t,x,y+\pi^-,d+\delta^-) - \tilde{v}^{(\lambda)}(t,x,y,d)\big] \;=\; 0 \\ \tilde{v}^{(\lambda)}(T,x,y,d) \;=\; \frac{1}{2} r(\eta,\beta)(d-x)^2. \end{cases} \tag{4.64}$$

We look again for a candidate solution to (4.64) in the form

$$\begin{aligned} \tilde{w}^{(\lambda)}(t,x,y,d) \;=\;& A_\lambda(T-t)(d-x)^2 + B_\lambda(T-t)y^2 + F_\lambda(T-t)(d-x)y \\ &+ G_\lambda(T-t)(d-x) + H_\lambda(T-t)y + K_\lambda(T-t), \end{aligned} \tag{4.65}$$

for some deterministic functions A_λ, B_λ, F_λ, G_λ, H_λ and K_λ. Plugging the candidate function $\tilde{w}^{(\lambda)}$ into equation (4.64), we see that $\tilde{w}^{(\lambda)}$ is solution to the HJB equation iff the following system of ordinary differential equations (ODEs) is satisfied by A_λ, B_λ, F_λ, G_λ, H_λ and K_λ :

$$\begin{cases} A'_\lambda + \frac{1}{4\gamma}(-2A_\lambda + \nu F_\lambda)^2 \;=\; 0 \\ B'_\lambda + \frac{1}{4\gamma}(2\nu B_\lambda - F_\lambda + 1)^2 \;=\; 0 \\ F'_\lambda + \frac{1}{2\gamma}(-2A_\lambda + \nu F_\lambda)(2\nu B_\lambda - F_\lambda + 1) \;=\; 0 \\ G'_\lambda - 2\mu A_\lambda + \frac{1}{2\gamma}(-2A_\lambda + \nu F_\lambda)(-G_\lambda + \nu H_\lambda) - \lambda(2\delta A_\lambda + \pi F_\lambda) \;=\; 0 \\ H'_\lambda - \mu F_\lambda + \frac{1}{2\gamma}(2\nu B_\lambda - F_\lambda + 1)(-G_\lambda + \nu H_\lambda) - \lambda(2\pi B_\lambda + \delta F_\lambda) \;=\; 0 \\ K'_\lambda - \mu G_\lambda - (\sigma_0^2 B_\lambda + \sigma_d^2 A_\lambda + \rho\sigma_0\sigma_d F_\lambda) + \frac{1}{4\gamma}(-G_\lambda + \nu H_\lambda)^2 \\ \quad -\lambda[(p^+(\delta^+)^2 + p^-(\delta^-)^2)A_\lambda + (p^+(\pi^+)^2 + p^-(\pi^-)^2)B_\lambda \\ \quad +(p^+\delta^+\pi^+ + p^-\delta^-\pi^-)F_\lambda + \delta G_\lambda + \pi H_\lambda] \;=\; 0 \end{cases}$$

with the initial conditions $A_\lambda(0) = \frac{1}{2}r(\eta,\beta)$, $B_\lambda(0) = 0$, $F_\lambda(0) = 0$, $G_\lambda(0) = 0$, $H_\lambda(0) = 0$, $K_\lambda(0) = 0$. We first solve the Riccati system relative to the triple $(A_\lambda, B_\lambda, F_\lambda)$, which is the same as in the no jump case, and therefore obtain: $A_\lambda = A$, $B_\lambda = B$, $F_\lambda = F$ as in (4.61). Then we solve the first-order linear system of ODE relative to the pair (G_λ, H_λ), which involves the jump parameters λ, π and δ, and get:

$$
\begin{aligned}
G_\lambda(t) &= G(t) + \frac{\lambda}{2}\frac{r(\eta,\beta)t(\pi t + 2\delta(\nu t + 2\gamma))}{(r(\eta,\beta)+\nu)t + 2\gamma}, \\
H_\lambda(t) &= H(t) - \frac{\lambda}{2}\frac{(\pi - 2r(\eta,\beta)\delta)t^2}{(r(\eta,\beta)+\nu)t + 2\gamma},
\end{aligned}
$$

where G and H are given from the no jump case (4.62). Finally, after some tedious but straightforward calculations, we explicitly obtain K_λ from the last equation:

$$
\begin{aligned}
K_\lambda(t) &= K(t) + \lambda\gamma\frac{p^+(\pi^+ - r(\eta,\beta)\delta^+)^2 + p^-(\pi^- - r(\eta,\beta)\delta^-)^2}{\big(r(\eta,\beta)+\nu\big)^2}\ln\Big(1 + \frac{(r(\eta,\beta)+\nu)t}{2\gamma}\Big) \\
&\quad - \frac{\lambda}{2}\frac{p^+((\pi^+)^2 - r(\eta,\beta)\delta^+(2\pi^+ + \nu\delta^+)) + p^-((\pi^-)^2 - r(\eta,\beta)\delta^-(2\pi^- + \nu\delta^-))}{r(\eta,\beta)+\nu}t \\
&\quad + \frac{\lambda r(\eta,\beta)}{2}\frac{2\nu\mu\delta + \lambda((p^+)^2\delta^+(\pi^+ + \nu\delta^+) + (p^-)^2\delta^-(\pi^- + \nu\delta^-))}{r(\eta,\beta)+\nu}t^2 \\
&\quad + \lambda^2\gamma r(\eta,\beta)\frac{r(\eta,\beta)\delta^2 + 2\nu p^+p^-\delta^+\delta^- - ((p^+)^2\delta^+\pi^+ + (p^-)^2\delta^-\pi^-)}{(r(\eta,\beta)+\nu)\big((r(\eta,\beta)+\nu)t + 2\gamma\big)}t^2 \\
&\quad + \frac{2\lambda\gamma r(\eta,\beta)^2\mu\delta}{(r(\eta,\beta)+\nu)\big((r(\eta,\beta)+\nu)t + 2\gamma\big)}t^2 - \frac{\lambda^2\pi^2}{48\gamma}t^3 \\
&\quad + \frac{\lambda^2 p^+p^- r(\eta,\beta)}{2}\frac{2\nu\delta^+\delta^- + \delta^-\pi^+ + \delta^+\pi^-}{(r(\eta,\beta)+\nu)t + 2\gamma}t^3 \\
&\quad + \frac{1}{8}\frac{4r(\eta,\beta)\mu\lambda\pi - \lambda^2\pi^2}{(r(\eta,\beta)+\nu)t + 2\gamma}t^3,
\end{aligned}
$$

with K in (4.63). The function $\tilde{w}^{(\lambda)}$ in (4.65) may thus be rewritten as the sum of $\tilde{w}$ in (4.60) and another function of t, $d - x$ and y, and is by construction a smooth solution with quadratic growth condition to the HJB equation (4.64). Moreover, the argmin in HJB equation for $\tilde{w}^{(\lambda)}$ is attained for

$$
\begin{aligned}
\tilde{q}^{(\lambda)}(t,x,y,d) &= -\frac{1}{2\gamma}\Big[\frac{\partial\tilde{w}^{(\lambda)}}{\partial x} + \nu\frac{\partial\tilde{w}^{(\lambda)}}{\partial y} + y\Big] \\
&= \frac{r(\eta,\beta)(\mu(T-t) + d - x) - y}{(r(\eta,\beta)+\nu)(T-t) + 2\gamma} \\
&\qquad + \lambda\frac{r(\eta,\beta)\delta(T-t) + \frac{\pi}{4\gamma}(r(\eta,\beta)+\nu)(T-t)^2}{(r(\eta,\beta)+\nu)(T-t) + 2\gamma} \\
&=: \hat{q}^{(\lambda)}(T-t, d-x, y).
\end{aligned}
$$

Again, notice that $\hat{q}^{(\lambda)}$ is linear, and Lipschitz in x, y, d, uniformly in time t, and so given an initial state (x, y, d) at time t, there exists a unique solution $(\hat{X}^{t,x,y,d}, \hat{Y}^{t,x,y,d}, D^{t,d})_{t\leq s\leq T}$ to (4.1)-(4.37)-(4.35) with the feedback control $\hat{q}_s^{(\lambda)} = \hat{q}^{(\lambda)}(T-s, D_s^{t,d} - \hat{X}_s^{t,x,y,d}, \hat{Y}_s^{t,x,y,d})$, which satisfies: $\mathbb{E}[\sup_{t\leq s\leq T} |\hat{X}_s^{t,x,y,d}|^2 + |\hat{Y}_s^{t,x,y,d}|^2 + |D_s^{t,d}|^2] < \infty$, see e.g. Theorem 1.19 in [64]. This implies that $\mathbb{E}[\int_t^T |\hat{q}_s^{(\lambda)}|^2 ds] < \infty$, hence $\hat{q}^{(\lambda)} \in \mathcal{A}_t$. We now call on a classical verification theorem for stochastic control of jump-diffusion processes (see e.g. Theorem 3.1 in [64]), which shows that $\tilde{w}^{(\lambda)}$ is indeed equal to the value function $\tilde{v}^{(\lambda)}$, and $\hat{q}^{(\lambda)}$ is an optimal control. Finally, once the optimal trading rate $\hat{q}^{(\lambda)}$ is determined, the optimal production is obtained from the optimization over $\xi \in \mathbb{R}$ of the terminal cost $C(D_T^{t,d} - \hat{X}_T^{t,x,y,d}, \xi)$, hence given by: $\hat{\xi}_T^{(\lambda)} = \frac{\eta}{\eta+\beta}(D_T^{t,d} - \hat{X}_T^{t,x,y,d})$. $\square$

Bibliographie

[1] René Aïd : *Electricity derivatives*. Springer Briefs in Quantitative Finance. Springer, Cham, 2015.

[2] René Aïd, Pierre Gruet et Huyên Pham : An optimal trading problem in intraday electricity markets. *Mathematics and Financial Economics*, 2015.

[3] Yacine Aït-Sahalia : Nonparametric pricing of interest rate derivative securities. *Econometrica*, 64(3) :527–560, 1996.

[4] Yacine Aït-Sahalia, Per A. Mykland et Lan Zhang : How often to sample a continuous-time process in the presence of market microstructure noise. *The Review of Financial Studies*, 18(2) :351–416, 2005.

[5] Yacine Aït-Sahalia, Per A. Mykland et Lan Zhang : Ultra high frequency volatility estimation with dependent microstructure noise. *Journal of Econometrics*, 160(1) :160–175, 2011.

[6] Aurélien Alfonsi, Antje Fruth et Alexander Schied : Constrained portfolio liquidation in a limit order book model. *In Advances in mathematics of finance*, volume 83 de *Banach Center Publ.*, pages 9–25. Polish Acad. Sci. Inst. Math., Warsaw, 2008.

[7] Aurélien Alfonsi, Antje Fruth et Alexander Schied : Optimal execution strategies in limit order books with general shape functions. *Quantitative Finance*, 10(2) :143–157, 2010.

[8] Aurélien Alfonsi et Alexander Schied : Optimal trade execution and absence of price manipulations in limit order books models. *SIAM Journal of Financial Mathematics*, 1(1) :490–522, 2010.

[9] Aurélien Alfonsi, Alexander Schied et Alla Slynko : Order book resilience, price manipulation, and the positive portfolio problem. *SIAM Journal of Financial Mathematics*, 3(1) :511–533, 2012.

[10] Robert Almgren : Optimal execution with nonlinear impact functions and trading-enhanced risk. *Applied Mathematical Finance*, 10(1) :1–18, 2003.

[11] Robert Almgren : Optimal trading with stochastic liquidity and volatility. *SIAM Journal on Financial Mathematics*, 3(1) :163–181, 2012.

[12] Robert Almgren et Neil Chriss : Optimal execution of portfolio transactions. *Journal of Risk*, 3 :5–39, 2000.

[13] Randolf ALTMEYER et Markus BIBINGER : Functional stable limit theorems for quasi-efficient spectral covolatility estimators. *Stochastic Processes and their Applications*, 2015.

[14] Federico M. BANDI et Jeffrey R. RUSSELL : Separating microstructure noise from volatility. *Journal of Financial Economics*, 79(3) :655 – 692, 2006.

[15] Alain BENSOUSSAN, Pierre BERTRAND et Alexandre BROUSTE : A generalized linear model approach to seasonal aspects of wind speed modeling. *J. Appl. Stat.*, 41(8) :1694–1707, 2014.

[16] Fred Espen BENTH et Steen KOEKEBAKKER : Stochastic modeling of financial electricity contracts. *Energy Economics*, 30 :1116–1157, 2008.

[17] Dimitris BERTSIMAS et Andrew W. LO : Optimal control of execution costs. *Journal of Financial Markets*, 1(1) :1–50, 1998.

[18] Ramaprasad BHAR et Carl CHIARELLA : Interest rate futures : estimation of volatility parameters in an arbitrage-free framework. *Applied Mathematical Finance*, 4 :181–199, 1997.

[19] Ramaprasad BHAR, Carl CHIARELLA et Thuy-Duong TÔ : A maximum likelihood approach to estimation of Heath-Jarrow-Morton models. Rapport technique 80, Quantitative Finance Research Centre, University of Technology, Sydney, 2002.

[20] Markus BIBINGER, Nikolaus HAUTSCH, Peter MALEC et Markus REIß : Estimating the quadratic covariation matrix from noisy observations : local method of moments and efficiency. *The Annals of Statistics*, 42(4) :80–114, 2014.

[21] Markus BIBINGER et Markus REIß : Spectral estimation of covolatility from noisy observations using local weights. *Scandinavian Journal of Statistics. Theory and Applications*, 41(1) :23–50, 2014.

[22] Petter BJERKSUND, Heine RASMUSSEN et Gunnar STENSLAND : Valuation and risk management in the norwegian electricity market. *In* Endre BJØRNDAL, Mette BJØRNDAL, Panos M. PARDALOS et Mikael RÖNNQVIST, éditeurs : *Energy, Natural Ressources and Environmental Economics*, pages 167–185. Springer, 2010.

[23] Ningyuan CHEN, Steven KOU et Chun WANG : A partitioning algorithm for markov decision processes and its application to limit order books with stochastic market depth. Working paper, 2013.

[24] Emmanuelle CLÉMENT, Sylvain DELATTRE et Arnaud GLOTER : An infinite dimensional convolution theorem with applications to the efficient estimation of the integrated volatility. *Stochastic Processes and their Applications*, 123(7) :2500–2521, 2013.

[25] Rama CONT, Arseniy KUKANOV et Sasha STOIKOV : The price impact of order book events. *Journal of Financial Econometrics*, 12(1) :47–88, 2013.

[26] John C. COX, Jonathan E. INGERSOLL JR. et Stephen A. ROSS : A theory of the term structure of interest rates. *Econometrica*, 53(2) :pp. 385–407, 1985.

[27] Sylvain DELATTRE et Jean JACOD : A central limit theorem for normalized functions of the increments of a diffusion process, in the presence of round-off errors. *Bernoulli*, 3(1) :1–28, 1997.

[28] Aryeh DVORETZKY : Asymptotic normality for sums of dependent random variables. *In* Lucien LE CAM, Jerzy NEYMAN et Elizabeth L. SCOTT, éditeurs : *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, pages 513–535. University of California Press, 1972.

[29] Peter A. FORSYTH, J. Shannon. KENNEDY, Shu Tong TSE et Heath WINDCLIFF : Optimal trade execution : a mean quadratic variation approach. *Journal of Economic Dynamics & Control*, 36(12) :1971–1991, 2012.

[30] Ernesto GARNIER et Reinhard MADLENER : Balancing forecast errors in continuous-trade intraday markets. FCN Working Papers 2/2014, E.ON Energy Research Center, Future Energy Consumer Needs and Behavior (FCN), 2014.

[31] Jim GATHERAL : No-dynamic-arbitrage and market impact. *Quantitative Finance*, 10(7) :749–759, 2010.

[32] Jim GATHERAL, Alexander SCHIED et Alla SLYNKO : Exponential resilience and decay of market impact. *In Econophysics of order-driven markets*, New Econ. Windows, pages 225–236. Springer, Milan, 2011.

[33] Valentine GENON-CATALOT et Jean JACOD : On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Annales de l'I.H.P.*, 29(1) :119–151, 1993.

[34] Valentine GENON-CATALOT, Catherine LAREDO et Dominique PICARD : Nonparametric estimation of the diffusion coefficient by wavelets methods. *Scandinavian Journal of Statistics*, 19(4) :317–335, 1992.

[35] Gregor GIEBEL, George KARINIOTAKIS, Richard BROWNSWORD, Caroline DRAXL et Michael DENHARD : The state-of-the-art in short-term prediction of wind power : A literature overview. Rapport technique, Deliverable Report D1.2 of the Anemos Project, 2011.

[36] Arnaud GLOTER et Jean JACOD : Diffusions with measurement errors. I. Local asymptotic normality. *ESAIM Probab. Statist.*, 5 :225–242, 2001.

[37] Arnaud GLOTER et Jean JACOD : Diffusions with measurement errors. II. Optimal estimators. *ESAIM Probab. Statist.*, 5 :243–260, 2001.

[38] Olivier GUÉANT, Charles-Albert LEHALLE et Joaquin FERNANDEZ-TAPIA : Optimal portfolio liquidation with limit orders. *SIAM J. Financial Math.*, 3(1) :740–764, 2012.

[39] Wolfgang HÄRDLE, Gerard KERKYACHARIAN, Dominique PICARD et Alexandre B. TSYBAKOV : *Wavelets, approximation, and statistical applications*, volume 129 de *Lecture Notes in Statistics*. Springer-Verlag, New York, 1998.

[40] David HEATH, Robert JARROW et Andrew MORTON : Bond pricing and the term structure of interest rates : A new methodology for contingent clams valuation. *Econometrica*, 60(1) :77–105, 1992.

[41] Arthur HENRIOT : Market design with centralized wind power management : handling low-predictability in intraday markets. *The Energy Journal*, 35(1) :99–117, 2014.

[42] Juri HINZ, Lutz VON GRAFENSTEIN, Michel VERSCHUERE et Martina WILHELM : Pricing electricity risk by interest rate methods. *Quantitative Finance*, 5(1) :49–60, 2005.

[43] Marc HOFFMANN : Minimax estimation of the diffusion coefficient through irregular samplings. *Statistics & Probability Letters*, 32(1) :11–24, 1997.

[44] Marc HOFFMANN : Adaptive estimation in diffusion processes. *Stochastic Processes and their Applications*, 79(1) :135–163, 1999.

[45] Marc HOFFMANN : L_p estimation of the diffusion coefficient. *Bernoulli*, 5(3) :447–481, 1999.

[46] Marc HOFFMANN, Axel MUNK et Johannes SCHMIDT-HIEBER : Adaptive wavelet estimation of the diffusion coefficient under additive error measurements. *Annales de l'Institut Henri Poincaré Probabilités et Statistiques*, 48(4) :1186–1216, 2012.

[47] Gur HUBERMAN et Werner STANZL : Price manipulation and quasi-arbitrage. *Econometrica*, 72(4) :1247–1275, 2004.

[48] Jean JACOD : On continuous conditional Gaussian martingales and stable convergence in law. *In Séminaire de Probabilités, XXXI*, volume 1655 de *Lecture Notes in Math.*, pages 232–246. Springer, Berlin, 1997.

[49] Jean JACOD : Statistics and high-frequency data. *In* Mathieu KESSLER, Alexander LINDNER et Michael SØRENSEN, éditeurs : *Statistical Methods for Stochastic Differential Equations*, pages 191–310. Chapman and Hall/CRC, 2012.

[50] Jean JACOD, Yingying LI, Per A. MYKLAND, Mark PODOLSKIJ et Mathias VETTER : Microstructure noise in the continuous case : the pre-averaging approach. *Stochastic Processes and their Applications*, 119(7) :2249–2276, 2009.

[51] Jean JACOD et Per A. MYKLAND : Microstructure noise in the continuous case : approximate efficiency of the adaptive pre-averaging method. *Stochastic Processes and their Applications*, 125(8) :2910–2936, 2015.

[52] Jean JACOD et Philip PROTTER : *Discretization of processes*, volume 67 de *Stochastic Modelling and Applied Probability.* Springer, Heidelberg, 2012.

[53] Jean JACOD et Mathieu ROSENBAUM : Quarticity and other functionals of volatility : efficient estimation. *The Annals of Statistics*, 41(3) :1462–1484, 2013.

[54] Jean JACOD et Albert N. SHIRYAEV : *Limit theorems for stochastic processes*, volume 288 de *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, deuxième édition, 2003.

[55] Andrew JEFFREY, Oliver LINTON, Thong NGUYEN et Peter C.B. PHILLIPS : Nonparametric estimation of a multifactor Heath-Jarrow-Morton model : An integrated approach. Rapport technique, Cowles Foundation for Research in Economics, 2001.

[56] Jussi Keppo, Nicolas Audet, Pirja Heiskanen et Iivo Vehviläinen : Modeling electricity forward curve dynamics in the nordic market. *In* D.W. Bunn, éditeur : *Modelling Prices in Competitive Electricity Markets*, pages 251–264. Wiley Series in Financial Economics, 2004.

[57] Rüdiger Kiesel, Gero Schindlmayr et Reik H. Börger : A two-factor model for the electricity forward market. *Quantitative Finance*, 9(3) :279–287, 2009.

[58] Steen Koekebakker et Fridthjof Ollmar : Forward curve dynamics in the nordic electricity market. *Managerial Finance*, 31(6) :73–94, 2005.

[59] Axel Munk et Johannes Schmidt-Hieber : Lower bounds for volatility estimation in microstructure noise models. *In Borrowing strength : theory powering applications—a Festschrift for Lawrence D. Brown*, volume 6 de *Inst. Math. Stat. Collect.*, pages 43–55. Inst. Math. Statist., Beachwood, OH, 2010.

[60] Axel Munk et Johannes Schmidt-Hieber : Nonparametric estimation of the volatility function in a high-frequency model corrupted by noise. *Electronic Journal of Statistics*, 4 :781–821, 2010.

[61] Marek Musiela et Marek Rutkowski : *Martingale methods in financial modelling*, volume 36 de *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin, deuxième édition, 2005.

[62] Per Mykland et Lan Zhang : The econometrics of high-frequency. *In* Mathieu Kessler, Alexander Lindner et Michael Sørensen, éditeurs : *Statistical Methods for Stochastic Differential Equations*, pages 109–190. Chapman and Hall/CRC, 2012.

[63] Anna A. Obizhaeva et Jiang Wang : Optimal trading strategy and supply/demand dynamics. *Journal of Financial Markets*, 16(1) :1–32, 2013.

[64] Bernt Øksendal et Agnès Sulem-Bialobroda : *Applied stochastic control of jump diffusions*. Universitext. Springer, Berlin, second edition, 2007.

[65] Huyên Pham : *Continuous-time stochastic control and optimization with financial applications*. Springer-Verlag, Berlin, 2009.

[66] Mark Podolskij et Mathias Vetter : Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli*, 15(3) :634–658, 2009.

[67] Markus Reiß : Asymptotic equivalence for inference on the volatility from noisy observations. *The Annals of Statistics*, 39(2) :772–802, 2011.

[68] Alfréd Rényi : On stable sequences of events. *Sankhyā (Statistics), Series A*, 25 :293–302, 1963.

[69] Mathieu Rosenbaum : Integrated volatility and round-off error. *Bernoulli*, 15(3) :687–720, 2009.

[70] Alexander Schied et Torsten Schöneborn : Risk aversion and the dynamics of optimal liquidation strategies in illiquid markets. *Finance and Stochastics*, 13(2) :181–204, 2009.

[71] Alexander SCHIED et Alla SLYNKO : Some mathematical aspects of market impact modeling. *In Surveys in stochastic processes*, EMS Ser. Congr. Rep., pages 153–179. Eur. Math. Soc., Zürich, 2011.

[72] Richard STANTON : A nonparametric model of term structure dynamics and the market price of interest rate risk. *The Journal of Finance*, 52(5) :1973–2002, 1997.

[73] Aad VAN DER VAART : *Asymptotic Statistics*. University of Cambridge, 1998.

[74] Alexander WEISS : Executing large orders in a microscopic market model. ArXiv e-prints, 2009.

[75] David WILLIAMS : *Probability with martingales*. Cambridge mathematical textbooks, 1991.

[76] Lan ZHANG : Efficient estimation of stochastic volatility using noisy observations : A multi-scale approach. *Bernoulli*, 12(6) :1019–1043, 2006.

[77] Lan ZHANG, Per A. MYKLAND et Yacine AÏT-SAHALIA : A tale of two time scales : Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100(472) :1394–1411, 2005.